

STIC-Biotech/ChemLib

138413

From: Chan, Christina
Sent: Monday, November 22, 2004 10:30 AM
To: Walicka, Malgorzata
Cc: STIC-Biotech/ChemLib
Subject: RE:

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Walicka, Malgorzata
Sent: Sunday, November 21, 2004 5:31 PM
To: Chan, Christina
Cc: STIC-Biotech/ChemLib
Subject:

Please authorize the RUSH search of SEQ ID NO:4 in the application No. 10/802,682. Please also against pending data.

Thank you.

Malgorzata A. Walicka, Ph.D.
Patent Examiner
Art Unit 1652, Recombinant Enzymes
USPTO, Remsen Building, Room 2C76
400 Dulany St.
Alexandria, VA 22313
Mail Room 2C70
Tel. (571)272-0944, fax (571)273-0944

RECEIVED
NOV 22 2004
STIC/ChemLib Division

STAFF USE ONLY

Searcher: Arnold
Searcher Phone: 2-2532
Date Searcher Picked up: 11/22/04
Date Completed: 11/23/04
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search

NA Sequence: # 1
AA Sequence: # _____
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIS: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: _____
WWW/Internet: _____
Other(Specify): _____

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 13:16:30 / Search time 7584 Seconds

(without alignments)
10849.694 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1740
1 atgaacccacacacgctgct.....tcttgactgcgcagctaa 1740

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database:

GenEmbl:.*
1: gb_ba:.*
2: gb_hlg:.*
3: gb_in:.*
4: gb_com:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_ses:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vl:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1740	100.0	1740	6	A70009 Sequence 4
2	1740	100.0	1740	6	CQ797619 Sequence 1
3	1130.2	65.0	1740	6	A70006 Sequence 1
4	1079.8	62.1	1737	6	A70008 Sequence 3
5	1053.6	60.6	1740	6	A70007 Sequence 2
6	625.8	36.0	22913	1	AB046580 Pseudoglu
7	141.8	8.1	299910	1	AP005957 Bradyrhiz
8	118.2	6.8	299910	1	AP005957 Bradyrhiz
9	110.2	6.3	300600	1	AP005936 Bradyrhiz
10	110.2	6.3	2097	1	AF277373 Ralstonia
11	101.2	5.8	2959	1	CTQHEH X81880 C.lescocter
12	98.8	5.7	7097	1	AF176640 Pseudomon
13	98.4	5.7	302101	1	AE016784 Pseudomon
14	95.2	5.5	110000	1	AY316747 Rhizobium
15	94.8	5.4	2350	1	AF355798 Pseudomon
16	94.4	5.4	4534	1	AB091400 Pseudomon
17	92.4	5.3	300200	1	AP005962 Bradyrhiz
18	88.8	5.1	2214	6	E03223 The gene en
19	88.8	5.1	4230	1	ABCADHCC D00635 Acetobacter

20	87.2	5.0	2802	1	AEADRGNE X82894 A.europaeus
21	87.2	5.0	4059	1	AEY09480 Y09480 A.europaeus
22	85.6	4.9	2163	6	AR143029 AR143029 Sequence
23	85.6	4.9	2163	6	BD269976 BD269976 Gluconoba
24	85.6	4.9	2265	6	AR143020 AR143020 Sequence
25	85.6	4.9	2265	6	BD269967 BD269967 Gluconoba
26	85.6	4.9	4830	6	AR143023 AR143023 Sequence
27	85.6	4.9	4830	6	BD269970 BD269970 Gluconoba
28	81.2	4.7	2272	1	AY048591 AY048591 Pseudomon
29	80.8	4.6	5602	1	D86375 D86375 Gluconobact
30	78	4.5	302101	1	AE016784 AE016784 Pseudomon
31	77.2	4.4	1694	1	PDECTT553T PDECTT553T Paracoccus
32	77.2	4.4	9470	1	PD034346 PD034346 Paracoccus
33	77	4.4	4915	1	PAE9858 PAE9858 Pseudomon
34	77	4.4	1533	1	AE004624 AE004624 Pseudomon
35	76.6	4.4	349635	1	BK572603 BK572603 Rhodospir
36	74.8	4.3	1869	1	AB162821 AB162821 Pseudomon
37	74	4.3	1989	1	AF326086 AF326086 Pseudomon
38	70.4	4.0	2625	9	AX079058 AX079058 Sequence
39	63	3.6	125020	9	AF429315 AF429315 Homo sapi
40	60.4	3.5	145153	1	AY458649 AY458649 Uncultured
41	60.4	3.5	2314	1	PDEM0H M17339 P.denitrifi
42	60.2	3.5	125020	9	AF429315 AF429315 Homo sapi
43	60	3.4	2011	1	AF309488 AF309488 Simorhizo
44	60	3.4	110000	1	RVE51985_01 RVE51985_01 Continuation (2 of
45	59.8	3.4	2466	1	MTBXXDH M2625 Methylobact

ALIGNMENTS

RESULT 1
LOCUS A70009 Sequence 4 from Patent EP0832974.
DEFINITION A70009
ACCESSION A70009
VERSION A70009.1 GI:4774465
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
HOFMANN LA ROCHÉ (CH)
Patent: EP 0832974-A 4 01-Apr-1998;

FEATURES
source
Location/Qualifiers
1..1740

ORIGIN
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

Query Match 100.0%; Score 1740; DB 6; Length 1740;
Best Local Similarity 100.0%; Pred. No. 6.1e-309;
Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAACCCCAACGCTGCTTCGACACGCGCGCGGCTGCTATTGCTTACCGCGCCGCC	60
DB	1	ATGAACCCCAACGCTGCTTCGACACGCGCGCGGCTGCTATTGCTTACCGCGCCGCC	60
QY	61	GCATTTCGCGAGTAAACCGATTAACGATGAACCTGTCGGAACCGCGCGCTGTGAA	120
DB	61	GCATTTCGCGAGTAAACCGATTAACGATGAACCTGTCGGAACCGCGCGCTGTGAA	120
QY	121	TGGAATTAACGCGCGCAACCAAGAACTATGCGACCTCGCCCGACCGAGTACT	180
DB	121	TGGAATTAACGCGCGCAACCAAGAACTATGCGACCTCGCCCGACCGAGTACT	180
QY	181	GCCGACACGTTGGTCACTGCACTGATCGGCGCGCGGATGAGGCGGCGCGCTA	240
DB	181	GCCGACACGTTGGTCACTGCACTGATCGGCGCGCGGATGAGGCGGCGCGCTA	240

QY	241	CAGGTCACGCCGATGATCCAGATGAGGCGTGATGATCTCGGCAAAACCCCGGATGATG	300
Db	241	CAGGTCACGCCGATGATCCAGATGAGGCGTGATGATCTCGGCAAAACCCCGGATGATG	300
QY	301	CAGGCGCTGAGATGCGCAACAGGAGATCTGATCTGGGAAACAACCGCGCAATCTGCCGC	360
Db	301	CAGGCGCTGAGATGCGCAAAAGGAGATCTGATCTGGGAAACAACCGCGCAATCTGCCGC	360
QY	361	GTGCGCACGCTAAACGCCCAAGGCGACCGCAAGCGCGCGTCGCGCTTTACGCAAGC	420
Db	361	GTGCGCACGCTAAACGCCCAAGGCGACCGCAAGCGCGCGTCGCGCTTTACGCAAGC	420
QY	421	CTCTATTCAGGTCATGAGCAACCATCTGATCCGCTGATATGAGACGGGCAAGTC	480
Db	421	CTCTATTCAGGTCATGAGCAACCATCTGATCCGCTGATATGAGACGGGCAAGTC	480
QY	481	GTAATCGATGTCGAACTGGGATCGGGCGAACAAGCGTTGACCAAGTAACCAACGGGGCG	540
Db	481	GTAATCGATGTCGAACTGGGATCGGGCGAACAAGCGTTGACCAAGTAACCAACGGGGCG	540
QY	541	ATTGTCGCAATGAGGTCATCGTCGCGGGTTCACCTCCAAATTTGCGCTATGATGTC	600
Db	541	ATTGTCGCAATGAGGTCATCGTCGCGGGTTCACCTCCAAATTTGCGCTATGATGTC	600
QY	601	TTTATCTCGGGCAGATTCGCGACGGGTGAGAGACTGTGGCCAAACATTATCCG	660
Db	601	TTTATCTCGGGCAGATTCGCGACGGGTGAGAGACTGTGGCCAAACATTATCCG	660
QY	661	CAGCGCGGCGAAGAGGGTGAAGAGCTGGGGCAATGATTTGAGAGGCGCGTGATGAC	720
Db	661	CAGCGCGGCGAAGAGGGTGAAGAGCTGGGGCAATGATTTGAGAGGCGCGTGATGAC	720
QY	721	GCGCTGTGGGTCAGATCACTATGATCCGTGACGAACCTTGTGTTCTATGCTGAC	780
Db	721	GCGCTGTGGGTCAGATCACTATGATCCGTGACGAACCTTGTGTTCTATGCTGAC	780
QY	781	GCGCTGGGTCAGGCTCCGAAACCCAGGCGGACCGCGGCGGCAAGCTGATGAGAC	840
Db	781	GCGCTGGGTCAGGCTCCGAAACCCAGGCGGACCGCGGCGGCAAGCTGATGAGAC	840
QY	841	AACACCCGCTTGGCGTGCCTGCCACAAGGAGAGATTGTCCTGGCGTCACAGACCTG	900
Db	841	AACACCCGCTTGGCGTGCCTGCCACAAGGAGAGATTGTCCTGGCGTCACAGACCTG	900
QY	901	CCGCGCGACACTGGGACCCAGATGACGCTTGAGATGATGCTGCCAAGCTGATGTC	960
Db	901	CCGCGCGACACTGGGACCCAGATGACGCTTGAGATGATGCTGCCAAGCTGATGTC	960
QY	961	CAACCTCGGCGGATGAGAGGCTTGGCGCGCAATCAACCCCAATGCGGCAACGGGAG	1020
Db	961	CAACCTCGGCGGATGAGAGGCTTGGCGCGCAATCAACCCCAATGCGGCAACGGGAG	1020
QY	1021	CGCGCTGTGCTGACGGGTGCGCTTGCAGAACCGGACAGATGTGCTTTGATGCGGCC	1080
Db	1021	CGCGCTGTGCTGACGGGTGCGCTTGCAGAACCGGACAGATGTGCTTTGATGCGGCC	1080
QY	1081	TGGGCGGAATTCCTGTGGGCGCGGTATCAACATCAACCAATATGATGGCTCGATGAC	1140
Db	1081	TGGGCGGAATTCCTGTGGGCGCGGTATCAACATCAACCAATATGATGGCTCGATGAC	1140
QY	1141	GAGACCGGCTTGTGACGTTGACAGAGATGCGGTGTTGAAAGACTGGAATAT	1200
Db	1141	GAGACCGGCTTGTGACGTTGACAGAGATGCGGTGTTGAAAGACTGGAATAT	1200
QY	1201	GACGTCGCGGACCTTCTGGGTGGGCGGACGTGTGTACGCGGCACTAAACCGGAC	1260
Db	1201	GACGTCGCGGACCTTCTGGGTGGGCGGACGTGTGTACGCGGCACTAAACCGGAC	1260
QY	1261	ACCGGCAATTAATTTCTGCGCTGACCAATGCTCTGAGATATTAAGCGGTGATCAA	1320
Db	1261	ACCGGCAATTAATTTCTGCGCTGACCAATGCTCTGAGATATTAAGCGGTGATCAA	1320
QY	1321	GAGTTTACGCGCTGACGCTCTATTAACACACGAGCGGACCGCAAACTGCGCGCGCTTT	1380

Db	1321	GAGTTTAGCGGCTGACGTCATTATTAACAACGCGCGACCGGMAAACTTCGGCCGGGGCTTT	1380
QY	1381	GAATAATATGCGCCGCACTGACGCGCATATGATATCAGCACCCGGCGCCACTTTGTGTGGCG	1440
Db	1381	GAATAATATGCGCCGCACTGACGCGCATATGATATCAGCACCCGGCGCCACTTTGTGTGGCG	1440
QY	1441	GAGGCGCCCTGGGGGAGACTCTACTCGCCGGTTTTCGACGCGGAGGCGGTGTGATTTCAAC	1500
Db	1441	GAGGCGCCCTGGGGGAGACTCTACTCGCCGGTTTTCGACGCGGAGGCGGTGTGATTTCAAC	1500
QY	1501	GGCGGAGACCGACCGCTATTTTCGTGCTTCAGCAGGAAACCGCGAGACTTTTGTGGCAG	1560
Db	1501	GGCGGAGACCGACCGCTATTTTCGTGCTTCAGCAGGAAACCGCGAGACTTTTGTGGCAG	1560
QY	1561	GCCCGCTTTCGAGGAGTCGCGACGCGGGCAGGCGCATCAGATCAGATTCGATTCGAGCGCGTCA	1620
Db	1561	GCCCGCTTTCGAGGAGTCGCGACGCGGGCAGGCGCATCAGATCAGATTCGATTCGAGCGCGTCA	1620
QY	1621	TATATCGCATCGGTGGGGCGGTCTTGAAGCAATGGCAATGGAAACGCGCGCGCTGGCC	1680
Db	1621	TATATCGCATCGGTGGGGCGGTCTTGAAGCAATGGCAATGGAAACGCGCGCGCTGGCC	1680
QY	1681	GAGGCATTCGATTCGACCTCGGTGGGTAAATCGATCATGCTTTTGCACTGCCACAGTAA	1740
Db	1681	GAGGCATTCGATTCGACCTCGGTGGGTAAATCGATCATGCTTTTGCACTGCCACAGTAA	1740

LOCUS	CQ797619	1740 bp	DNA	Linear	PAT 20-APR-2004
DEFINITION	Sequence 1 from Patent WO2004029267.				
ACCESSION	CQ797619				
VERSION	CQ797619.1	GI:46425899			
KEYWORDS					
SOURCE					
ORGANISM	Glucobacter oxydans				
	Glucobacter oxydans				
	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;				
	Acetobacteraceae; Glucobacter.				
REFERENCE	1				
AUTHORS	Hoshino, T. and Shinjoh, M.				
TITLE	Process for producing L-ascorbic acid				
JOURNAL	Patent: WO 2004029267-A 1 08-APR-2004;				
	DSM IP Assets B.V. (NL)				
FEATURES					
source	location/qualifiers				
	1..1740				
	/organism="Glucobacter oxydans"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:442"				
CDS	1..1740				
	/note="unamed protein product"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAG26844.1"				
	/db_xref="GI:46425900"				
	/translation="MNPTTLRTSAVILLTAPAAFNQVPTIDELLANPAGWIVNY				
	GRNDENVYHSPILQITADNVQQLQVWARGNAGAVQVTPNIDGVIMLAVPGVIOQ				
	LDAQTGLIWEHRQLPAVATINAGDKRGKVALYGTSLFSSVDNHILADMTGVOQ				
	VEFDVGRSGEDGLSTNTGPIVANGVIVASGTQCYSPYGCPISSHDSATGEILRNHRH				
	IPOGREGGDEGTGNDPEARMTGTVGGOITVPVNLVFGSTGVGPASETORTGPIGGT				
	LYGNTRRAVAPDPRGEIYMRQTLPRNMMDCEPFEMVNAVDPQSAEMEGALAINA				
	NAAGDERVLTGAPCKTGITWMSFLPASGEFIMADTYNNILASIDPTGLVYNDAIN				
	LKEIDVERVLTGAPCKTGITWMSFLPASGEFIMADTYNNILASIDPTGLVYNDAIN				
	SATKTLAFAFENMRIDAIDISTGRITWMSARPLANTSPVISTAGGVFNQGTIDTFYF				
	ALSTQETGLTQARLATVATGOAISYELDGVQYIYAIGAGGLTYGTQNLNPLAEIDST				
	SVGNALVYFALPQ"				


```

Db      1 ATGAACCCCAACAGCTGCTTGGACACGCGCGCTGCTATTGCTTACCGGCGCCGCC 60
Qy      61 GCATTCGCGGAGGTAAACCCGATTTACCGATGAAGACTGTGGCGAACCGCGCGCTGTGAA 120
Db      61 GCATTCGCGGAGGTAAACCCGATTTACCGATGAAGACTGTGGCGAACCGCGCGCTGTGAA 120
Qy      121 TGAATTAACAGCGCGCAACCAAGAAATCTATCGCACTGGCGCCCTGACCCAGATCACT 180
Db      121 TGAATTAACAGCGCGCAACCAAGAAATCTATCGCACTGGCGCCCTGACCCAGATCACT 180
Qy      181 GCGGACAACTTGGTCACTGTCGAACTGTGTCGCGCGCGGAGTGAAGCGGCGGCGTA 240
Db      181 GCGGACAACTTGGTCACTGTCGAACTGTGTCGCGCGCGGAGTGAAGCGGCGGCGTA 240
Qy      241 CAGGTCACGCGGATGATGCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 300
Db      241 CAGGTCACGCGGATGATGCAATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCT 300
Qy      301 CAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Db      301 CAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Qy      361 GTGCGCAACGCTTAAACCGCGCAAGCGGACCGGACCGGCGCTGCTTAAACCGCGCA 420
Db      361 GTGCGCAACGCTTAAACCGCGCAAGCGGACCGGACCGGCGCTGCTTAAACCGCGCA 420
Qy      421 CTCTATTTCAGCTCATGAGCAACCACTGATGCTGATGCTGATGATGAGAGCGGCGCAGT 480
Db      421 CTCTATTTCAGCTCATGAGCAACCACTGATGCTGATGCTGATGATGAGAGCGGCGCAGT 480
Qy      481 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db      481 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Qy      541 ATATGCGCCCAATGAGCGTCACTGTCGCGGCTTCCACCTGCGCAATATGCGCTGATGATG 600
Db      541 ATATGCGCCCAATGAGCGTCACTGTCGCGGCTTCCACCTGCGCAATATGCGCTGATGATG 600
Qy      541 ATATGCGCCCAATGAGCGTCACTGTCGCGGCTTCCACCTGCGCAATATGCGCTGATGATG 600
Db      541 ATATGCGCCCAATGAGCGTCACTGTCGCGGCTTCCACCTGCGCAATATGCGCTGATGATG 600
Qy      601 TTTATCTCGGCGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCACTTTATCCG 660
Db      601 TTTATCTCGGCGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCACTTTATCCG 660
Qy      661 CAGCGCGGCGAAGAGGTGACGAGCTTGGCGGCAATGATTTGAGAGCGCGCTGATGATGAC 720
Db      661 CAGCGCGGCGAAGAGGTGACGAGCTTGGCGGCAATGATTTGAGAGCGCGCTGATGATGAC 720
Qy      721 GAGGCTGAGGAGTCAATCACTATGATCCGATGACGAACTTGTGTTCTATGAGCTGAC 780
Db      721 GAGGCTGAGGAGTCAATCACTATGATCCGATGATCCGATGACGAACTTGTGTTCTATGAGCTGAC 780
Qy      781 GAGGCTGAGGAGTCAATCACTATGATCCGATGATCCGATGATCCGATGATCCGATGAT 840
Db      781 GAGGCTGAGGAGTCAATCACTATGATCCGATGATCCGATGATCCGATGATCCGATGAT 840
Qy      841 AACACCCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Db      841 AACACCCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Qy      841 AACACCCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Db      841 AACACCCGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 900
Qy      901 CCGCGGACAACTGGGACCAAGATGACGATGATGATGATGATGATGATGATGATGATG 960
Db      901 CCGCGGACAACTGGGACCAAGATGACGATGATGATGATGATGATGATGATGATGATG 960
Qy      961 CAACCCCTCGGCGAGATGAGAGGTCTGCGCGCATGACCCCAATGCGGCGGACGCGGCGAG 1020
Db      961 CAACCCCTCGGCGAGATGAGAGGTCTGCGCGCATGACCCCAATGCGGCGGACGCGGCGAG 1020
Qy      1021 CCGCGTGTGCTGACGCGGTGCGCTTGCAGAACCGGACGAGTGTGCTGTTGATGCGGCG 1080
Db      1021 CCGCGTGTGCTGACGCGGTGCGCTTGCAGAACCGGACGAGTGTGCTGTTGATGCGGCG 1080
Qy      1081 TCGGCGGAATTCCTGTGGCGGCTGATACCACTAACCAATATGATGCCCTGCATGCAGC 1140

```

```

Db      1081 TCGGCGGAATTCCTGTGGCGGCTGATACCACTACCAATATATGCTCCATGCAGC 1140
Qy      1141 GAGACCGGCTTGTGACGCTGTAACGAGATGCGGTGCTGAAAGCTGAGACCTTGAAATAT 1200
Db      1141 GAGACCGGCTTGTGACGCTGTAACGAGATGCGGTGCTGAAAGCTGAGACCTTGAAATAT 1200
Qy      1201 GAGCTCTCGCGACCTTCTGCGGTGGCGCGGACTGCTGATGATGATGATGATGATGATG 1260
Db      1201 GAGCTCTCGCGACCTTCTGCGGTGGCGCGGACTGCTGATGATGATGATGATGATGATG 1260
Qy      1261 ACCGCGATTTACTTCTTTCGCGCTGAAACATGCTGCTGATGATGATGATGATGATGAT 1320
Db      1261 ACCGCGATTTACTTCTTTCGCGCTGAAACATGCTGCTGATGATGATGATGATGATGAT 1320
Qy      1321 GAGTTTACGCGGCTGAGCTGATTAACACGAGCGCGACCGGAAACTGCGCGCGCTTT 1380
Db      1321 GAGTTTACGCGGCTGAGCTGATTAACACGAGCGCGACCGGAAACTGCGCGCGCTTT 1380
Qy      1381 GAAATATAGGCGCGCATGACGCGATGATATACAGACCGGCGCACCTTGTGCTGCG 1440
Db      1381 GAAATATAGGCGCGCATGACGCGATGATATACAGACCGGCGCACCTTGTGCTGCG 1440
Qy      1441 GAGCGCCCTTGGCGGGAATTAATCTGCGCTTTTGTGACAGCGAGCGGTGTGCTTCAAC 1500
Db      1441 GAGCGCCCTTGGCGGGAATTAATCTGCGCTTTTGTGACAGCGAGCGGTGTGCTTCAAC 1500
Qy      1501 GCGCGGACCGACCGCTATTTCTGCTGCGCTCAGCGAGGAAACCGGCGGAGCTTTGCGAG 1560
Db      1501 GCGCGGACCGACCGCTATTTCTGCTGCGCTCAGCGAGGAAACCGGCGGAGCTTTGCGAG 1560
Qy      1561 GCGGCTTGTGAGCGGTGCGACGCGGCGGAGCTGACCTGATGATGATGATGATGATGATG 1620
Db      1561 GCGGCTTGTGAGCGGTGCGACGCGGCGGAGCTGACCTGATGATGATGATGATGATGATG 1620
Qy      1621 TATATGCGCATGCGGTGCGGCGGCTGACCTGATGATGATGATGATGATGATGATGATG 1680
Db      1621 TATATGCGCATGCGGTGCGGCGGCTGACCTGATGATGATGATGATGATGATGATGATG 1680
Qy      1681 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
Db      1681 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740

RESULT 3
A70006 1740 bp DNA linear PAT 07-MAY-1999
LOCUS A70006
DEFINITION Sequence 1 from Patent EP0832974.
ACCESSION A70006
VERSION A70006.1 GI:4774462
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1740)
AUTHORS Tomiyama,N., Asakura,A., Ojima,S., Hoshino,T. and Shinjoh,M.
TITLE Alcohol-aldehyde-dehydrogenases
JOURNAL Patent: EP 0832974-A.1 01-Apr-1998;
HOFPMANN LA ROCHE (CH)
FEATURES
source
1..1740
Location/Qualifiers
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN
Query Match 65.0%; Score 1130.2; DB 6; Length 1740;
Best Local Similarity 78.8%; Pred. No. 3.6e-197;
Matches 1374; Conservative 0; Mismatches 363; Indels 6; Gaps 2;
Qy      1 ATGAACCCCAACAGCTGCTTGGACACGCGCGCTGCTATTGCTTACCGGCGCCGCC 60
Db      1 ATGAACCCCAACAGCTGCTTGGACACGCGCGCTGCTATTGCTTACCGGCGCCGCC 60

```

```

QY 61 GCATTGCGGAGTAAACCCGATTAACGATGATGCGGAGAACCCGCGCTGGTAA 120
Db 61 GCGTTGCTCAAGTGAACCCCGCTCAACGATGATGCGGAGAACCCGCGCTGGTAA 120
QY 121 TGGATTAACTAGGCGCGCAACCAAGAAAATAATGCGACTGCGCCCTGACCAATCACT 180
Db 121 TGGATTAACTAGGCGCGCAACCAAGAAAATAATGCGACTGCGCCCTGACCAATCACT 180
QY 181 GCGGACAAAGTTGTGTGATGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
Db 181 ACTGAACAAAGTTGTGTGATGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
QY 241 CAGGTGACGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Db 241 CAGGTGACGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 CAGGCGCTGGATGCGCAACAGCGCATGATGATGATGATGATGATGATGATGATG 360
Db 301 CAGGCGCTGGATGCGCAACAGCGCATGATGATGATGATGATGATGATGATGATG 360
QY 361 GTCGACAGCTAAACGCGCAAGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 361 ATGCGACGCTGAACAGCTTTGGCGAGCGGACCGCGCGATGCGCGTGAACGCAAC 420
QY 421 CTCTATTGAGCTCATGCGAACAACATCTGATGCGCTGATGATGATGATGATGATG 480
Db 421 GTTTACTTTGTTTCTGCGGACAAACCACTGCTGCTGACACCGCACTGCGCAAGT 480
QY 481 GTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
Db 481 AGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
QY 541 ATTGTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 541 ATGTGTCGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
QY 601 TTTATCTCGGAGGACGATTTCCGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAG 660
Db 601 TTTATCTCGGAGGACGATTTCCGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 CAGCGCGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
Db 661 CAGCGCGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 GCGGTCTGGGAGTCAATCACTTAATGATCCCGTGAACCACTTGTGTTATAGCTGAC 780
Db 721 GCGGTCTGGGAGTCAATCACTTAATGATCCCGTGAACCACTTGTGTTATAGCTGAC 780
QY 781 GGGGTGGGCGCGAGTCCGAAACCAAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 840
Db 781 GGGGTGGGCGCGAGTCCGAAACCAAGCGGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 841 AACACCGCGCTTTGCGGTGCGTCCGACAGCGGCGAGATTTGCTGCGCTGAACCA 900
Db 841 AACACCGCGCTTTGCGGTGCGTCCGACAGCGGCGAGATTTGCTGCGCTGAACCA 900
QY 901 CCGCGCGCAACTGCGGAGCAAGAGATGATGATGATGATGATGATGATGATGATG 960
Db 901 CCGCGCGCAACTGCGGAGCAAGAGATGATGATGATGATGATGATGATGATGATG 960
QY 961 CAACCCCTGCGCGAGATGAGAGGTCTGCGCGCATCAACCCCAATGCGCGCGAG 1020
Db 961 CAACCCCTGCGCGAGATGAGAGGTCTGCGCGCATCAACCCCAATGCGCGCGAG 1020
QY 1021 CCGCGGTGCTGAGCGGTGCGCTTGAAGAAGCGGACAGATGATGATGATGATGATG 1080
Db 1021 CCGCGGTGCTGAGCGGTGCGCTTGAAGAAGCGGACAGATGATGATGATGATGATG 1080
QY 1081 TCGGCGCAATTTCTGTGCGCGCTGATACCACTGAACAACAATGATGATGATGATG 1140
Db 1081 TCGGCGCAATTTCTGTGCGCGCTGATACCACTGAACAACAATGATGATGATGATG 1140
QY 1141 GAGACGCGGCTTTGATGAGGTGAACGAGAGATGCGGTGCTGAAGAAGCTGGAATAT 1200

```

```

Db 1138 GAAACGGCATGCTGACCGGTGAACGAAGATGCGATCTGAAGAACTGATGTGAATAT 1197
QY 1201 GAGCTGCGCGGACCTTCCGCGGTGCGGAGATGCTGCGACCGCATGAACCGGAC 1260
Db 1198 GAGCTGCGCGGACCTTCCGCGGTGCGGAGATGCTGCGACCGCATGAACCGGAC 1257
QY 1261 ACCGCGATTTACTTCTTGGCGCTGAACAAATGCTGCTAGCAATTAATGACCGTTG 1320
Db 1258 AGCGGATCTACTTCACTCCCGTGAACAAAGTCTGCTATGACATGATGCGCGATG 1317
QY 1321 GAGTTTAGCGCGCTGACCGTGAACCAACAGCGGACCGCAAAATCGCGCGGCTTT 1380
Db 1318 GAATTCACCTGATGAGCGGTGAACCAACAGCAAGTGAACAGTGTGCGCGGCAAG 1377
QY 1381 GAAATATGCGCGCATGACGCGATGATATGACGACCGCGGCGCACTTGTGTGCGG 1440
Db 1378 GATATGATGCGGTGATGATGACGATGACATGACGACGCGGTGATGCGGTGTGCTG 1437
QY 1441 GAGCGCGCTGCGCGGAACTACTGCGCGTTTGTGCGAGCGGAGCGGTGTGCTGAC 1500
Db 1438 GAACGTGCTGCGCGGAACTACTGCGCGTTTGTGCGAGCGGAGCGGTGTGCTGAC 1497
QY 1501 GCGCGGACCGACCGCTATTTCGCTGCGCTGACGCAAGAAACCGCGGAGACTTTGTG 1560
Db 1498 GGTGTGACGATGCTTACTTCCCGCGCTTCAAGCAAGAAACCGCGGAGACTTTGTG 1557
QY 1561 GCGCGCTTGCAGCGGTGCGGACGCGGAGCGGATGACTGATGATGATGATGATG 1620
Db 1558 ACCGCGCTTCAACCGCGGTGCGGAGCGGATGACTGATGATGATGATGATGATG 1617
QY 1621 TATATGCGCAATGCGGTGCGGAGCGGATGACTGATGATGATGATGATGATGATG 1679
Db 1618 TATGTGCGCAATGCGGATGCTGATGATGATGATGATGATGATGATGATGATG 1677
QY 1680 --CGAGCAATGCAATTTGACCTGCGTGTGATGCAATGATGATGATGATGATG 1737
Db 1678 GCGGAGCGAGTCACTGACCGCATGCGTGAACGCGGTGATGATGATGATGATGATG 1737
QY 1738 TAA 1740
Db 1738 TAA 1740

```

RESULT 4

```

LOCUS A70008 1737 bp DNA linear PAT 07-MAY-1999
DEFINITION Sequence 3 from Patent EP0832974.
ACCESSION A70008
VERSION A70008.1 GI:4774464
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1737)
AUTHORS Tomiyama, N., Asakura, A., Ojima, S., Hoshino, T. and Shinjoh, M.
TITLE Alcohol-aldenhyd-dehydrogenases
JOURNAL Patent: EP 0832974-A 3 01-Apr-1998;
HOFMANN IA ROCHE (CH)
FEATURES
location/Qualifiers
1..1737
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

```

ORIGIN

```

Query Match 62.1%; Score 1079.8; DB 6; Length 1737;
Best Local Similarity 77.1%; Pred. No. 6,2e-188;
Matches 1341; Conservative 0; Mismatches 392; Indels 6; Gaps 2;
QY 1 ATGAACCCCAAGCTGCTTGGACCAAGCGGCGCTGCTATTTACCGGCGCGCC 60
Db 1 ATGAACCTGACGACCTGCTGCAAGCAAGCGGCGCTGCTGCTTGGACCAATTC 60

```

```

QY 61 GCATTGCGGAGGTAACCCGATTAACCGATGAATGCTGTGGCGAACCCTGCGCTGGTGA 120
Db 61 GCGCTTGCCCA--AACCGCATACCGATGAATGCTGGCGAACCCTGCGCTGGTGA 117
QY 121 TGAATTAATACGCGCGCAACCAAGAAATATGCGCACTGCGCCCTGACCCAGATCACT 180
Db 118 TGAATCAATACGCTGCAAGCAAGAGACTAACCGCACTGCGCCCTGACCCAGATTAAC 177
QY 181 GCGCAACAGTGGTCACTGTCGCACTGTCTGGCGCGCGGAGATGAGAGCGGGGCGGTA 240
Db 178 GCAACACAGCTGGCGCACTGCACTGTCTGGCGCGCGGATGAGAGCGGCAAGATC 237
QY 241 CAGGTACGCGAGATTCATGATGAGGTGATGATGCAAAACCCCGGTGATGATG 300
Db 238 CAGGTACCCCGCTGTGTCATGACGGGTGATGATGCAAAACCCCGGTGATGATG 297
QY 301 CAGGCGCTGAGATGCGCAACAGCGGATCTGATCTGGGACACCGCGCCAACTGCGCGC 360
Db 298 CAGGCGCATGACCGCGGACCGGCGATCTGATCTGGGACACCGCGCCAACTGCGCGA 357
QY 361 GTGCGCACGCTAAACGCCCAAGCGACCGGAGCGGCGGCTGCGCCCTTAACGCGACG 420
Db 358 ATGCGCACGCTGAACAGCTTTGTGAGCCGACCGCGGCGGCTGATGCGACGAC 417
QY 421 CTCTATTGAGCTCATGAGGACAAACCATCTGATCGCGCTGATATGAGACGCGCCAGTTC 480
Db 418 GTCTATTGCTCTGCTGGGACCAACTGTGTGCGGTGAGACCTGACCGCGCCAAAGTC 477
QY 481 GTATTGATGATGCACTGATGAGTGGGGGAGAGCGGCTTGACCACTGACACCAAGCGGCG 540
Db 478 GTATTGACGCTGATGCGGCTGCAAGGACCGGATA--TGTCTGCAACTGTCGCGCGCG 534
QY 541 ATTGTCGCATGAGCGCTCATGCTGCGGAGTTCCACCTGCAATATTGCGCCCTATGATG 600
Db 535 ATTGTCGCATGAGCGCTCATGCTGCGGAGTTCCACCTGCAATATTGCGCCCTATGATG 594
QY 601 TTATATCTCGGGGCAAGATCCGCGAGCGGATGAGAGCTGTGGCGCAACCACTTATCCCG 660
Db 595 TTGCTTTCGGGCGACGATCGGCGCACCGGATGAGAGCTGTGGCGCAACCACTTATCCCG 654
QY 661 CAGCGCGGCGAAGAGGATGAGAGCTGTGGGCGAATGATTGAGGCGGCGCTGGATGAC 720
Db 655 CCGCGCGGCGAAGAGGATGAGAGCTGTGGGCGAATGATTGAGGCGGCGCTGGATGAC 714
QY 721 GCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
Db 715 GCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 774
QY 781 GCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
Db 775 GCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 834
QY 841 AACACCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 835 AACACCGGCTTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 894
QY 901 CCGCGGCACTGCGGACCAAGATGAGCTTGAATGATGATGATGATGATGATGATGATGATG 960
Db 895 CCGCGGCACTGCGGACCAAGATGAGCTTGAATGATGATGATGATGATGATGATGATGATG 954
QY 961 CAGACCTGCGGATGAGAGGATGAGAGCTGCGGCGCATGACCCCAATGCGGCGAGCGGCG 1020
Db 955 CAGACCTGCGGATGAGAGGATGAGAGCTGCGGCGCATGACCCCAATGCGGCGAGCGGCG 1014
QY 1021 CCGCGGCTGAGTGAAGGATGAGAGCTGCGGCGCATGACCCCAATGCGGCGAGCGGCG 1080
Db 1015 CCGCGGCTGAGTGAAGGATGAGAGCTGCGGCGCATGACCCCAATGCGGCGAGCGGCG 1074
QY 1081 TCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
Db 1075 ACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1134

```

```

QY 1141 GAGACCGGCTTGTGACGCTGACGAGAGATGGGCTGCTGAAAGACTTGGAATAT 1200
Db 1135 CCGGATGGTCTGTGTGATGAGAGAGATCTGTGTGAGACCGAGCTGGAAGTGGCTTAT 1194
QY 1201 GAGCTCTGCGCGGACCTTCTGAGGATGAGAGCTGAGTGTGAGCGGACCTGAAACCGGAC 1260
Db 1195 GAAATCTGCGGACCTTCTGAGGATGAGAGCTGAGTGTGAGCGGACCTGAAACCGGAT 1254
QY 1261 ACCGCGATTTACTTCTTGGCGCTGAAACATGCTGCTACATATTATGAGCGGTTGATCA 1320
Db 1255 ACTGCGATCTATTTCATCCCGTGAACACCGCTGTACCGGTATGACGCTGTGACCA 1314
QY 1321 GAGTTAGGCGGCTGAGAGTCTATTAACACAGCGGACCGGAAACCTGCGCGGCGCTT 1380
Db 1315 GAGTTAGGCGGCTGAGAGTCTATTAACAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 1374
QY 1381 GAAATATGAGGCGGATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATG 1440
Db 1375 GAAATATGAGGCGGATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATG 1434
QY 1441 GAGGCGGCTGAGGAGAACTACCTGCGGCTTGTGAGAGCGGCGGCTGATGATGATG 1500
Db 1435 GAGGCGGCTGAGGAGAACTACCTGCGGCTTGTGAGAGCGGCGGCTGATGATGATG 1494
QY 1501 GCGGCGGACGACCGGCTATTTCCGTCCTGAGCGGAAACCGGCGAGACTTGTGCGAG 1560
Db 1495 GCGGCGGACGACCGGCTATTTCCGTCCTGAGCGGAAACCGGCGAGACTTGTGCGAG 1554
QY 1561 GCGGCTTGTGAGAGCTGAGAGCGGCGGAGGATGAGCTGAGCTGAGCTGAGCTGAG 1620
Db 1555 ACCGCTTGTGAGAGCTGAGAGCGGCGGAGGATGAGCTGAGCTGAGCTGAGCTGAG 1614
QY 1621 TATATGCGGATGAGGCGGCGGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1680
Db 1615 TATATGCGGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1674
QY 1681 GAGGCGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1739
Db 1675 GAGGCGATGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 1733

```

RESULT 5

LOCUS	A70007	1740 bp	DNA	linear	PAT 07-MAY-1999
DEFINITION	Sequence 2 from Patent EP0832974.				
ACCESSION	A70007				
VERSION	A70007.1 GI:4774463				
KEYWORDS	unidentified				
SOURCE	unidentified				
ORGANISM	unclassified				
REFERENCE	1 (bases 1 to 1740)				
AUTHORS	Tomiyama, N., Asakura, A., Ojima, S., Hoshino, T. and Shinjoh, M.				
TITLE	Alcohol-aldehyde-dehydrogenases				
JOURNAL	Patent: JP 0832974-A 2 01-Apr-1998;				
FEATURES	HOFMANN LA ROCHER (CH)				
source	Location/Qualifiers				
ORIGIN	1..1740				
Query Match	/organism="unidentified"				
Best Local Similarity	/mol_type="unassigned DNA"				
Matches 1118; Conservative 0; Mismatches 419; Indels 3; Gaps 1;	/db_xref="taxon:32644"				

```

QY 1 ATGAACCCCAACAGCTGCTTCCGACACGAGCGGCGGATATGCTTACCGGCGCGCC 60
Db 1 ATGAACAGCTGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
QY 61 GCATTGCGGAGGTAACCCGATTAACCGATGAATGCTGTGGCGAACCCTGCGCTGGTGA 120

```

Db 61 GGGCTTGTCTAAGTGAACCCCGCTCAACGATGATTTGTGGGAGAACCCCGCGCTGTGTAA 120
 Qy 121 TGGATTAACTACGGCGGCAACCAAGAAAATATCGGCACTCGCCCTTGAACCAATCACT 180
 Db 121 TGGATTACGCTACGGCTAGAGAACCAAGAAAATATCGGCACTCGCCCTTGAACCAATCACT 180
 Qy 181 GCGGCAAGAGTTGGTCAATGCACTGGCTGGGCGCGGAGATGAGAGGCGGCGCTA 240
 Db 181 ACTGAGAAAGTGGCCCACTGCACTGGCTGGGCGCGGAGATGAGAGGCGGCGCTA 240
 Qy 241 CAGGTACGCGCGATGATCCATGATGCGGTATCTGGCAAAACCCCGGTATGTATC 300
 Db 241 CAGGTACGCGCGGTATGATGATGCGGTATCTGGCAAAACCCCGGTATGTATC 300
 Qy 301 CAGGCGCTGGATGCGCAACAGCGGATCTGATCTGGGAAACAACCGCGCAACTGCGCGC 360
 Db 301 CAGGCGCTGGATGCGCAACAGCGGATCTGATCTGGGAAACAACCGCGCAACTGCGCGC 360
 Qy 361 GTGCGACGCTAAACCGCCCAAGCGCAACCGCGCGGCTGCGCTTTACGCGACGAGC 420
 Db 361 ATGCGACGCTGAACAGCTTTGGCGAGCGACCGCGGCGCATGGCGCTGACGCGACCAAC 420
 Qy 421 CTCTATTTCACTCATGAGCAACCATCTGATCGCGCTGATATGAGACGCGGCGAGTC 480
 Db 421 GTTTACTTTGTTCTGTGGGCAACCACTGTGTGCTTGGACACGCACTGGCAAGTG 480
 Qy 481 GTATTGATGTGCAAGTGGATCGGCGGAAAGACGCGCTTGAACGATTAACAACAGCGGCGC 540
 Db 481 AGCTTGAAGTGAACCGCGCGCAAGCGGCGGCAAGACATGCTTTCG---AAGTGTGGGCGCG 537
 Qy 541 ATTGTGCGAATGCGCTCATCTGTGCGGCTTCACTTGCATATTTGCGCTCATATGATG 600
 Db 541 ATTGTGCGAATGCGCTCATCTGTGCGGCTTCACTTGCATATTTGCGCTCATATGATG 600
 Qy 538 ATGTGTGCAAAACGCGCTGATCTGTGCGGCTTCACTTGCATATTTGCGCTCATATGATG 597
 Db 538 ATGTGTGCAAAACGCGCTGATCTGTGCGGCTTCACTTGCATATTTGCGCTCATATGATG 597
 Qy 601 TTTATCTGGGCGGCAAGATCCGCGACGCGGAGAGCTGTGGCGCAACCACTTTATCCG 660
 Db 598 TTTGTCTGGGCGGCAAGATCCGCGACGCGGAGAGCTGTGGCGCAACCACTTTATCCG 657
 Qy 661 CAGCGCGGCGAAGAGGTGAACAGACTTGGGCGGATGATTTGAGGCGCGCTGTGATGAC 720
 Db 658 CCGGCTGGCGAAGAGGTGATGAGACTTGGGCGGCAAGATTTAGAGACCGCTGTGATGAC 717
 Qy 721 GGGCTCTGGGCGGATGATCAACCTATGATCCGCGGACCAACCTGTTGATGATGATGATG 780
 Db 718 GGGCTCTGGGCGGATGATCAACCTATGATCCGCGGACCAACCTGTTGATGATGATGATG 777
 Qy 781 GGGCTGGGCGGACGCTGCCAACCAGCGCGGACCGCGGCGGACCGCTGTATGAGCAC 840
 Db 778 GCTGTGGCGCGGCTTCCGAAACCCAGCGCGGACCAACCGCGGCGGACCATGTATGAGCAC 837
 Qy 841 AAGACCGGCTTTGGCGGTGCTCCGACAGCGGCGGATGTTCTGGGCTGACAGACCGCTG 900
 Db 838 AAGACCGGCTTTGGCGGTGCTCCGACAGCGGCGGATGTTCTGGGCTGACAGACCGCTG 897
 Qy 901 CCGCGCGAACAAGTGGGCAAGATCAAGTTCGATGATGATGATGATGATGATGATGATG 960
 Db 898 CCGCGCGAACAAGTGGGCAAGATCAAGTTCGATGATGATGATGATGATGATGATGATG 957
 Qy 961 CAACCTCGGCGGAGATGAGGCTGTGCGCGCATCAACCCCAATTCGCGCGACGCGCGAG 1020
 Db 958 CAGCGCGGAGATGAGGCTGTGCGCGCATCAACCCCAATTCGCGCGACGCGCGAG 1017
 Qy 1021 CGCGGTGCTGACGCGGTGCGCTTGCAGACCGGCGGATGATGATGATGATGATGATG 1080
 Db 1018 CGTGGCGTTTGAACCGGCGTTCGCTGCAAAACCGGCGCATGTGCGAGTTGACGCGGAA 1077
 Qy 1081 TCGGCGGAATTCGTGTGGGCGGCTGATACCACTAACAATATGATGCTGCTGATGAC 1140
 Db 1078 ACCGGCGAATTCGTGTGGGCGGCTGATACCACTAACAATATGATGCTGCTGATGAC 1137
 Qy 1141 GAGACCGGCGCTTGTGACGCGTGAACAGGATCGGCTGTGAAAGAGCTGAGCTGATGAT 1200
 Db 1138 GAGACCGGCGCTTGTGACGCGTGAACAGGATCGAAGATTCGATGAGCTGAGACCGCTTAT 1197

Qy 1201 GAGCTGACCCGACCTTCCCTGGGTGGGCGCGACTGTGCTGACCGCACTGAAACCGGAC 1260
 Db 1198 GAGCTGACCCGCTGCTGCTGGTGGCGGATGAGCTGGCCGCTGCGCTGAAACCCGAT 1257
 Qy 1261 ACCGGCATTTACTTCTTGGCGCTGAACAATGCTGCTGATGATATGATGATGATGAT 1320
 Db 1258 ACCGGCATTTACTTCTTGGCGCTGAACAATGCTGCTGATGATATGATGATGATGAT 1317
 Qy 1321 GAGTTTACGCGCTGACGCTGATTAACACAGCGGACCGCAAACTCGCGCGGCTT 1380
 Db 1318 GAATTCAGCTGCTGACGCTGATTAACACAGCGGACCGCAAACTCGCGCGGCTT 1377
 Qy 1381 GAAATATGCGCGCGCACTGACGCGATGATATGACGACCGGCGGACCTTGTGTGAGG 1440
 Db 1378 GAGCTGCTGCGCGTATGACGCGCATGACATGACGACGCGGCGACCTGTGTGAGG 1437
 Qy 1441 GAGCGCGCTGCGCGGCACTGACGCGCTTGTGACGCGGCGGCGGCTGTGTGAGG 1500
 Db 1438 GAGCGCGGACCTGCGGCACTGACGCGCTTGTGACGCGGCGGCGGCTGTGTGAGG 1497
 Qy 1501 GCGCGGACCGGCTATTTCTGCTGACGCGGCGGCAAAACCGGCGGACCTTGTGTGAGG 1560
 Db 1498 GCGCGGACCGGCTATTTCTGCTGACGCGGCGGCAAAACCGGCGGACCTTGTGTGAGG 1557
 Qy 1561 GCGCGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620
 Db 1558 ACCGCTGTGCGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1617
 Qy 1621 TATATGCGCATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1680
 Db 1618 TATATGCGCATGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1677
 Qy 1681 GAGCGCATGATGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 Db 1678 GAGCGCGTGAATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1737

RESULT 6
 AB046580
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 ORGANISM
 Pseudomonas aeruginosa
 REFERENCES
 1. Shibata, T. and Saito, Y.
 Alcohol dehydrogenase
 JOURNAL
 TITLUS
 REFERENCES
 2. (bases 1 to 2913)
 Submitted (26-JUL-2000) Takashi Shibata, Fujisawa Pharmaceutical
 Co., Ltd., Fermentation Development Laboratories, Shinkawa-cho
 Nakagawa 156, Nishikawaguni, Aichi 462-0915, Japan
 (E-mail: takashi.shibata@fujisawa.co.jp, Tel: 81-52-401-0275,
 Fax: 81-52-400-1380)
 FEATURES
 source
 1. 2913
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="IFO 14464"
 /db_xref="taxon:133921"
 209..2035
 /gene="adh"
 209..2035
 /gene="adh"
 /codon_start=1
 /transl_table=1

```
/product="alcohol dehydrogenase"  
/protein_id="BAB62258.1"  
/db_xref="GI:15022439"  
/translation="MRFEYLROVNVGLALSTALISLSEPAFOHDANAABPSXAG  
SALENROPYADLAKNPANPILRGVQGWGYSPLDQINDQNGDLOVSRTPMP  
GSNEGAAIAYNGVIFLGNTRDYQALDGTGSLINRYRKLPSAKFTINSLSGARST  
ALFGDVFVSWDNFVALDARTKGLAMFTINGQVEBSVANSGIYVDGVILGST  
CQPSGFCVYTGIDASGEELMRNFIPIRGSGDDTWGAGYENRMWTGAWGQITTD  
PELDIVYSGTSGAPASEVORGETEGTLAGTIRFAVEXKTEVVMKQITLFRDMS  
ECTFEWVSTSVNPAKADGMSVAGNVPGRGTAVLGVPCXGVAMOPDARKGDY  
FNKATVEONSLASIDTGLVYEMDMILKEPKTYNCPFTLGGRDMSAGTPEKSN  
LYIPLSNACVDMATTEATPADYNDYNTLVLAPGKMRVRVADLDAENETKSY  
ETRAALYDVLTTGGDILVFVGGIDRDLRLDASGEVWSTLPGHVSQITTSYIDG  
RQIVAVSGLSGFTFPPTPDVDSAGANGIYVFLPERK"
```

ORIGIN

```
Query Match 36.0%; Score 625.8; Db 1; Length 2913;  
Best Local Similarity 61.8%; Pred. No. 9,5e-105;  
Matches 1066; Conservative 0; Mismatches 647; Indels 12; Gaps 4;
```

```
24 CACCAAGCGGCGCTGCTATTGCTTACCGCGCCGCCGATTGCGCAGGTAACTCCGAT 83  
Db CAACGCGCGCTGCCGACCGAGCAAGCGCGCCAGAGCTATCGAATCTCCAGCCGCT 366  
Qy 84 TACCGATGAACTGCTGGCGAACCAGCGCGTGTGAATGGAATTAACTACGCGCGCAACA 143  
Db CAGGCTGACGATCTGCGCGCAAGAACCGCGCAACTGCGCGATCTGCGCGGTAACTA 426  
Qy 367 CAGGCTGACGATCTGCGCGCAAGAACCGCGCAACTGCGCGATCTGCGCGGTAACTA 426  
Db 144 AGAAAATATATGCGCACTGCGCCCTGAACCGAGATCACTGCGCAACAGCTTGCTAGTGA 203  
Qy 427 CAGGCGTGGGGTTACTGCGCGCTCGACGAGTCAACAAGAACAGTGGGCACTTCA 486  
Db 204 ACTGCTTGGGCGCGCGGAGTGAAGCGCGCGCGCTGACAGTCAAGCTCAAGCTATCATTA 263  
Qy 487 GCTGCTGCTGCGCGCAAGTGAAGCGCGCTCAACAAGAGTGTGCGCATGCGCTATTA 546  
Db 264 TGGCGTATGATCTGCGCAACCCCGGATGATGATCAGCGCGCTGATGCGCAACAG 323  
Qy 547 CCGGCTCATCTTCTGCGCAACCAACGATGATCAGCGCGATGAGCGCAAGCGG 606  
Db 324 CGATCTGATGCGGAACACCGCGCGCAACTGCGCGCGGTGCGCAAGCTTAAAGCGCCA 380  
Qy 607 TTCCCTCATCTGGAATATGTCGAAGCTGCGCGCTCCAACTTATCACTGCT 666  
Db 381 AGCGCAACCGCAAGCGCGCTGCGCGCTTAAAGCGCAAGAGCTTATTCAGCTCAATGGA 440  
Qy 667 CCGCGCGCGCAAGCTTCAATGCACTCTTCCGCGCAAGGTCTATTTGTGTGCTGGA 726  
Db 441 CAACCATCTGATCGCGCTGATGAGAACGCGCGCAAGTGTATTCGATGTCGAACGTGG 500  
Qy 727 CAACCTTGATGCTGCGCTCGATGCAAGACCGCGCAAGCTGCGTGGGAACCAATGCGG 786  
Db 501 ATGCGCGCAAGAGCGCTTGAACGATGACCAACGCGCGCGATGTCGCCAATGCGCTAT 560  
Qy 787 CCGGCGGCTTGAAGAGCGCTGCGCACTTCTCCGCGCGATGCTGATGAGTGGCGTGT 846  
Db 847 GATCGAGGCTTCAACCTGCGCACTTCTCGGCTTGGTGTATGTAACCGGTAACGACGC 906  
Qy 621 CCGGACGGGAGAGAGCTGTGCGCAACCACTTATCCCGCAACCGCGCGGAAGAGGTGA 680  
Db 907 TGAAGTCGGTGAAGAACTGTGCGCAACCTTCACTCCCGCTCGGCGGAAGAGGTGA 966  
Qy 681 CGAGACTTGGGCAATG---ATTTCAGGCGCGCTGATGACCGGCTGCGGCTGACAT 737  
Db 967 CGACACCTGGGGGGCGCGCCCTACGAAAACGTTGATGACCGGCTGCGGCGCAAGT 1026  
Qy 738 CACTTATGATCCCGTGAAGAACTTGTGTTCTATGCTGACCGGCGTGGCGCGCAAGCTTC 797  
Db 1027 CACTTATGATCCCGAAGCTGCACTGTGTTACTACGCTGACCGGCGCGCGCGCTTTC 1086  
Qy 798 CGAAACCGCAAGCGGCGCGCGCGCTGATGGAACCAACCGCTTTCGCT 857
```

```
Db 1087 GGAAGTCAGGCGCGCGACCGAAGCGCGCACTTCGACAGGACCAATCCCGCTTTCGCT 1146  
Qy 858 GCGTCCCGCAACCGGCGCGATTTGCTGCGCGCTCAACGACCCCTGCGCGCGCAACTGGGA 917  
Db 1147 GAAGCCCAAGACCGGTAAGTGTCTGGAAGACCAACCTGCGCGCGCAACTGGGA 1206  
Qy 918 CCAAGATGACGCTTCCGATGATGTCGCCAAGTGCATGTGCAACCTTCGCGCGAGAT 977  
Db 1207 CTCGGAATGACGCTTCCGATGATGTCGCCAAGTGCATGTGCAACCTTCGCGCGAGAT 1266  
Qy 978 GGAAGTCTGCGCGCGCACTCAACCCCAATGCGCGCGAGCGGA---GGCGCGTGTGAC 1034  
Db 1267 CGATGCAATGATGTCGCTGCGTCCAAAGTGCAGCGCGCGCAAGACCGCAAGTGTGAC 1326  
Qy 1035 GGGTCCGCTTGAAGACCGGCGAGATGTCGTTGATGTCGCGCTTCGCGCGAATTCCT 1094  
Db 1327 CCGGCGTCCGCGCAAGACCGGCGTTCGCTGCGCAATTCGATGCGCAAGACCGGCGACTT 1386  
Qy 1095 GTGCGCGGCTATACCACTTACCAATATGATGTCGCTTCGATGCGAGAGACCGGCTTGT 1154  
Db 1387 CTGCTCAAGGCAACGCTGCGAGAACTGATGCTTCATGATGACACAGGCTTCGT 1446  
Qy 1155 GACGATGACAGGATGCGGTCGTAAGAGCTGAGCTTGAATATGACGCTTCCCGAC 1214  
Db 1447 TACGCTCAATGAGACATGATCTCAAGAGACCGGCGAAGCTTACATGATGATGATG 1506  
Qy 1215 CTTCCTGCGGCGCGCGCGCTGTCGTCGTCAGCGCGCACTGAACCGCGCACTTATCTT 1274  
Db 1507 CTTCCTGCGGCGGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 1566  
Qy 1275 CTTCGCGGTCGCAATGCTGCTGTAAGATATGATGATGATGATGATGATGATGATGATG 1334  
Db 1567 GATCCCGCTCAGCAACCGCTCTGTCAGATGATGATGATGATGATGATGATGATGATGAT 1626  
Qy 1335 CAGCGCTATTAACCAACGCGCGCAACCTGCGCGCGGCTTTGAAAATATGAGCG 1394  
Db 1627 TGAAGCTTAAACCAACGCGCGCAACCTGCGCGCGGCTTTGAAAATATGAGCG 1686  
Qy 1395 CATGACCGGCTATTAATCAACCAACCGGCGCACTTTCGTCGCGCGGCGCGCGCG 1454  
Db 1687 GGTGATGCTCATGATCTGCGCAACCGGCGAGACCAAGTGTGCTTACAAACCGGCGGCG 1746  
Qy 1455 GAACTACTCGCGCGCTTTTGTGACGCGAGCGCGGTGTGTTCAACGCGCGGACGACG 1514  
Db 1747 TCTGTCAGACCGGCTTCGACCAACCGGCGGCGACCTGCTGTGCTGCGGTATGATG 1806  
Qy 1515 CTATTTCCGTCCTTCAACGAGAAACCGGCGAGACTTTGTGCGAGGCGGCTTTCGAC 1574  
Db 1807 TGACTTCCGCTCTGACGCGCGAGTCCGCAAGAGTCTGTCACCCGCTTCGCGG 1866  
Qy 1575 GGTGCGAGCGGCGAGGATCAGCTGAGTTGAGCGCGCGTGCATATATGCGCATCG 1634  
Db 1867 TCGGCTTCCGCTTACACCACTACTTCAATGATGATGATGATGATGATGATGATGATGAT 1926  
Qy 1635 TCGCGCGCTGACCTATGAGCAAGCAATGAAACGCGCGCTGCGCGAGCAATTC 1694  
Db 1927 CTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1983  
Qy 1695 GACCTCGGTCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1739  
Db 1984 GCTTTCGCGCGCAACGCGCATGATGATGATGATGATGATGATGATGATGATGATGATG 2028
```

```
RESULT 7  
AP005957  
LOCUS 29910 bp DNA linear BCT 12-SEP-2003  
DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section  
23/31.  
ACCESSION AP005957 BA000040  
VERSION AP005957.1 GI:27354278  
KEYWORDS Bradyrhizobium japonicum USDA 110  
SOURCE
```

ORGANISM Bradyrhizobium japonicum USDA 110
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobiium.

REFERENCE 1
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
Sasamoto,S., Watanabe,A., Idesawa,K., Iriiguchi,M., Kawashima,K.,
Kohata,M., Matsumoto,M., Shimo,S., Tsunoka,H., Wada,T., Yamada,M.
and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110

JOURNAL DNA Res. 9 (6), 189-197 (2002)

MEDLINE 22484998

PUBMED 12597275

AUTHORS 2
Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T.,
Sasamoto,S., Watanabe,A., Idesawa,K., Iriiguchi,M., Kawashima,K.,
Kohata,M., Matsumoto,M., Shimo,S., Tsunoka,H., Wada,T., Yamada,M.
and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium
Bradyrhizobium japonicum USDA110 (supplement)

JOURNAL DNA Res. 9 (6), 225-256 (2002)

MEDLINE 22485002

PUBMED 12597279

AUTHORS 3 (bases 1 to 299910)
Kaneko,T.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, 2-6-7
Kazusa-Kametari, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp/
URL: http://www.kazusa.or.jp/rhizobase/
Tel:81-438-52-3935 (ex.2338) Fax:81-438-52-3934)

FEATURES
source Location/Qualifiers
1..299910
/organism="Bradyrhizobium japonicum USDA 110"
/mol_type="genomic DNA"
/strain="USDA110"
/db_xref="taxon:224911"
7..1380
/gene="b1r6001"
7..1380
/note="ORF ID:b1r6001"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BACS1266.1"
/db_xref="GI:27354279"
/translation="MVDHFKRDLPGAIVMLGYLARQSVVALVPMYEQADVPRGLG
ALVNYARPVNLDLMSFAKAGIALVLDTEGQVLAEGKNSPEMAAAVKGSGYADI
LAGYFPGNRRLDAFLTAGTKRPEQLHGTGCFDPAARWMLADGEPQGLVNNAN
FPLVNGRFTGCGGEREAMVAVGAWDGAIVYRIDLKQVFAVLAIDRLAAAPDRIT
ILVRPPESEEDYRNALAHANVVIDGTSGVLDIRNAAVVHLNCGAAVSVLLGK
LPIOLEXYLPTTGAHALPARVSSVNSFDELLGADIDHIERETETPDPAVHADIE
APFHLDGQAERVADVLTDAANSRRPYSLATYKGRDKPSIQIYKGAASAVLGS
AVTERLRSRFPARDKRIPEVFTLLQRIAHDPASPAQFATRAHCATTGLPLAS
IAIEPAR"
1384..2787
/gene="b1r6002"
1384..2787
/note="ORF ID:b1r6002"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BACS1267.1"
/db_xref="GI:27354280"
/translation="MATSARATLITLTAAYOTRFMTI PVACRLRAGHVELLAEPDRS
AEKSAEGVPTNMRREGKAGSPDDRGAFFARVAAYGLDSTNLFSEHKEPTGIRD
TAAERRENTYANAEVTLDRBGHQAELVQEGGSPVLSASTFAARNTKRFI
EBSFPRGRMYFTPDITGADVMSPADVVSPPVRAVLDITLAVRAIVIPKDDHYS
ARKKVNLNRRRLTEKMDOPALGEGHGRHLRHARVHAMALNARLKLTPIP
DAFVYVPHVADMALTLRSPDYLDQVATVDFLRTITPDSHVLVKEHPAQIGAIISA

gene
CDS
1384..2787
/gene="b1r6002"
/note="ORF ID:b1r6002"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BACS1267.1"
/db_xref="GI:27354280"
/translation="MATSARATLITLTAAYOTRFMTI PVACRLRAGHVELLAEPDRS
AEKSAEGVPTNMRREGKAGSPDDRGAFFARVAAYGLDSTNLFSEHKEPTGIRD
TAAERRENTYANAEVTLDRBGHQAELVQEGGSPVLSASTFAARNTKRFI
EBSFPRGRMYFTPDITGADVMSPADVVSPPVRAVLDITLAVRAIVIPKDDHYS
ARKKVNLNRRRLTEKMDOPALGEGHGRHLRHARVHAMALNARLKLTPIP
DAFVYVPHVADMALTLRSPDYLDQVATVDFLRTITPDSHVLVKEHPAQIGAIISA

DRLEFARPEDFNVLLPQOTNNTVTLNRADAVTSVNSKGAELILGKVVYNGDAFY
RSCPLVYADVRLADVPAURLREALSAGPDPARGAPYFESAMSSSYGELIYISDPKLLD
TFPAAISRAAIAEAPRVN"
2799..3548
/gene="b1r6003"
2799..3548
/note="ORF ID:b1r6003"
probable glycosyl transferase
/codon_start=1
/transl_table=11
/protein_id="BACS1268.1"
/db_xref="GI:27354281"
/translation="MPLVSLITPSNVEGLIERTISVNSQTFADWELLIAADCSTDR
TPEIVADIGEDRRVKLIRKPKGFPALRQAHIDRQRTAFLDSDMLPGRKER
QLAPARKRALSLYTLARRINENTVTSGLISVASTTGQLKNTAIVLAVDRS
IACPIVKNKGNDYDFCLMLSLKRGHTAGLDLARYVRSVSSRVPVRAKAWQ
IYRNVEHLISVRSWCFPAHGAAMLKREEF"
3706..4668
/gene="b1r6004"
3706..4668
/note="ORF ID:b1r6004"
/codon_start=1
/transl_table=11
/product="dUDP-glucose 4-6-dehydratase"
/protein_id="BACS1269.1"
/db_xref="GI:27354282"
/translation="MPEESFKSRILIVTGAGFGISHTCERLIDGAEVYSDNFTG
SRNNIHLINLPLFEAVRDVTFPLIEVDALFNLCPSFPIHQSDPVQTKTSVHG
AIVMLGLAKLAKARIFQASTSEVYGDPLHPQEDWNGVNPYIGRSYDEKRCLET
LFEDYRQGLPIKVARIEFTYGPROMDGRVSSFIYQALQGEPIITFVGQGGQTRS
FCYVDLVERIKRMVTKEDITGPIVGNNSERTIELEKVIETLGTGSKLVFKPLP
ODDPRORQDPLTKAKTALMWERKVALDEDGLKETIAVFKSLSEIA"
4711..6024
/gene="b1r6005"
4711..6024
/note="ORF ID:b1r6005"
unknown protein
/codon_start=1
/transl_table=11
/protein_id="BACS1270.1"
/db_xref="GI:27354283"
/translation="MSAPSGIMQHLQSLDSKRALITLLLSAVTCLSLIKVAFQFS
YIHFGSRVWIAVAVAVAFSVSLFPAARFSFGAGFYFTYMLIGFIMDIVFSDYS
YDLRLAGLSAASLVLLPLFVRAPFQOLVLSNAGEHLITALLVSVGTIVAS
TVNRLVATINIVYRDALAPGAVRYLIGWVSTLLPPTFACYSLLGRMRAGVLI
LLFPYPTITTKARFETPAMLITLAVLSSELEARTSVTSIFVPMILGLVIAVGA
INSLPGRFDVLVIRMAITFSALDITNHFZANPITWFCQISMLKPNVPCAYODLS
VWENNTYFQMIVASIPATEGVASVGPYLAFLTALVAGFTLADGNFASGLPQRLVLI
SSQTLPHVLTNLEPLTVAMVHGTALLFLIMVWPSIFQPGLPD"
6021..7199
/gene="b116006"
6021..7199
/note="ORF ID:b116006"
probable glycosyl transferase"
/codon_start=1
/transl_table=11
/protein_id="BACS1271.1"
/db_xref="GI:27354284"
/translation="MPEAPRIATFTNIPRIWAGVYVORNLPAALHRSPTQVAV
LFGTRDDPEELALASTRAVALVRSPAIDRATGCLGALACRDPALAPPAOQVD
ITVESARFEGMRLPCPALAMPDFQRLPLPSSRAVYRSDGFPQALASGHVLS
SASALDLKRFPGISNGSVNFAFEPVSLATPSVIAQYGLPSYFIYINQFW
RHKHQVYVDALGLLKGRIIDVVAASGSTRDEPDLPERKMSVRSGLTENFHL
GMPLDHLVVALRASMAILNPSCEGMSITVEAKSGFVPMIISDIDVHREQTEGAR
YFGRIDAEALAAHLSVAQSDPPTIRDLPLDEEVKFAFVDFVLTADVMQNSGIR
"
7192..8124
/gene="gale"
7192..8124
/complement (7192..8124)


```
/gene="galE"  
/note="ORF ID: b116007  
probable UDP-glucose 4-epimerase (EC 5.1.3.2)"  
/transl_table=1  
/codon_start=1  
/protein_id="BAC51272.1"  
/db_xref="GI:27354285"  
/translation="MSAWTGANGFGRHLVRELAVGRTVHGSHGALDPAEARALG  
LOTINGEVDANLSALATHGPISOIFHLAGSSVGLSTARFEDPSFTVSTARLT  
EMLRSPRESELIVASAAVYGDHAGPIAESALAPMSPYGHKTKMQLCRSYAQS  
FGHCTVYRLFSYVGRYLRKQLMDTOSRLSSESLNIGTGAEIRDTVDVRL  
LVCAAEVQDEDFRLINGSGRGTSVAELIRWGSSTVYRSGVAPGDPASLT  
ADGRLDMNFDRLLPRLERGLADYEMFRSQAARA"  
/complement (8205 .8825)  
/gene="b116008"  
/complement (8205 .8825)  
/gene="b116009"  
/note="ORF ID: b116008  
putative acetyltransferase"  
/transl_table=1  
/codon_start=1  
/protein_id="BAC51273.1"  
/db_xref="GI:27354286"  
/translation="MSLSLTIYRNCVHAIDFNLATYVRLIGRATCRIDRGAFLHS  
ARLNALGDEKIVIGSHSHVRELMTLHGQGINIGWCYGVGRIMSASIDIGN  
RVLISHSVNIFDLNLSHPIRASERHEQAKELFTGHPROLSDLRPIKICDDAMIGACA  
MVRGVYIGEGGVAAGAVTVTKDVPAYSIVAGNPATIVRELSPDER"  
/complement (8903 .9991)  
/gene="b116009"  
/complement (8903 .9991)  
/note="ORF ID: b116009  
unknown protein"  
/transl_table=1  
/codon_start=1  
/protein_id="BAC51274.1"  
/db_xref="GI:27354287"  
/translation="MRLFQNSCFYPSYLPRLNOLAANKAVSFERRRVLNDRFGASHF  
LQVLVDSPEAFPTNGDDDEVLRMARERQMGEPFLLEILLAQLEHAETVEFLNDF  
VRYPSFAVALPGCKVKALRCGAPSGNADFTAYGVLANFPLMDMWRKCCRELFDP  
VPADHVMDEYGESEPIDIVFVGVGSRHREKATLEVARLADSHVEVFLDASRL  
TKLASAIGRLPLRQHRPDAIAIARPPFQDLYELIGSAKILNDSIMAGDR  
GNRCFESMGGALLVSDAGRTIDMESGITTEYTPRHAAEVLARSIDMPRSKEI  
AALGRARIDYHKMOKLFEVVGRL"  
/complement (9995 .11835)  
/gene="b116010"  
/complement (9995 .11835)  
CDS  
Query Match 8.1%; Score 141.8; DB 1; Length 29910;  
Best Local Similarity 47.3%; Pred. No. 2.9e-16;  
Matches 714; Conservative 0; Mismatches 737; Indels 60; Gaps 7;  
QY 125 TTAACGAGCCGCAACCAAGAAAATATGCACTGCGCCCTGACCCAGATCACTGCCG 184  
DB 224840 TCAACTACGGGATGGGCTAACAATCCAGCGCTTCGACCCGACCAAGATCAACAAG 224899  
QY 185 ACAACGTGGTCACTGGCACTGCTCGGGCCCGCGGATGAGCGGGGGCCGTACAG 244  
DB 224900 ACAACGTGAAAAACCTCGTCCCGGTGGAACTACTCTTCAATGACGATCGACGGAAG 224959  
QY 245 TCAACGCGATGATCCATGATGGGTGATGATCTGGCAAAACCCCGGTGATGTGATCCAG 304  
DB 224960 AATCGACGACACTGAGTGTACAGGGCGGTGATTAACGTAACCTGCAACAACGACCATG 225019  
QY 305 CGCTGATGGCAAAACAGGGGATCTGATCTGGGAAACACGCGCCCACTGCCCGCTG 364  
DB 225020 CGGTGACGCAAGACCGGCAACGATCTGGAATTCAGATCGATACCCCGCGGAGA 225079  
QY 365 CCAAGCTAAAGCCCAAGGACCG--CAAGCGCGCGCTGCGCTTTTACGCGACAGCC 421  
DB 225080 CGCGCGCATCGTCTGTGGGATCATCAACGCGCGCGCGCGCTCTAGACAGGCAAG 225139  
QY 422 TCTATTTCAGCTATGGGACAAACATCTGATCGCGCTGATATGAGACGGGCGAGGTG 481
```

```
DB 225140 TCTTCCGACCAACCTTCGACGCGCAACGTGATGCGCTCGACGCCACGAGGACGCAAGGAG 225199  
QY 482 TATTCGATGTGCAACGTGATCGGGGCGAAGACGGCTTGACCAATGACCAACGAGGCGCA 541  
DB 225200 TTGGCGGACAGAGCGCGCCCATATCAAGAGGCTATTTCATG--ACGGTGGCGCCG 225256  
QY 542 TTGTGGCAATGAGCGGTCACTCGCGGGGTTTCACTGCAATATTCGCGCTATGGATGCT 601  
DB 225257 TGGTGGCGACGGCGTCTGCTATCACCGGCATCTCGGCGCGGACGTTGCGACCGCGCT 225316  
QY 602 TTAATCTGGGAGCAATTCGCGACGCGTGAAGAGCTGTGCGCAACCACTTATCCGC 661  
DB 225317 TCAATGACGGCTGGGATCCGGGACAGGGCAAGCATCTCTGCGCACCACTTATCCCT 225376  
QY 662 AGCCCGGAGAGAGGTACGAGACTTGGGGCATGATTTGAGAGCGCGGTGATGACCG 721  
DB 225377 CTCCGACGAGCGCGGTGGCGACACTGGAAGGCGACACTGGAAGCTCGGCGGCT 225436  
QY 722 GCGTCTGGGTCAGATCACTATGATCCGTTGACGAACCTTGTGTTATGACTGACCG 781  
DB 225437 CGACTTGATACCGGCTCTTACGATCCCGAGTGAACAGGTCTATTGGGGCATCGGCA 225496  
QY 782 GCGTGGGCCCAAGCTGCGAAGCCAGCGGCGACCGCGGCGGACGCTGATGGCACCA 841  
DB 225497 ATCCCGGCGCTTCACCTCGGCGGTGCGC-----CCGCGCAACCTCTA---CACT 225547  
QY 842 ACACCGCTTTCGCGTGGCTGCGCGCACAGGCGAGATGTGTGGGTCACGACCGCTGC 901  
DB 225548 GCTCCGTGCTGCGATGATCCCAAGACCGGCAAGATCAATGGCACTACAGTTCTGCG 225607  
QY 902 CGCGGACACTGGGACCAAGATGACAGTTCGAGATGATGTCGCCAACGTGATGTC 961  
DB 225608 CGAACATCCGTTGCACTATGACAGCGGCGAGATGCTCGCGCAATGAACGTCG 225667  
QY 962 AACCTCGGCGAGATGAGAGGCTGTGCGCGGCAATCAACCCCAATCGGCGGACGCG 1021  
DB 225668 AGGCAACCCACCAAGGTGTGATGATGCAAC----- 225703  
QY 1022 GCGGTGCTGACGCGGTGCGCTTGCAAGACCGGACGATGATGCTTTGATGCGGCT 1081  
DB 225704 ----GTAAAGGCTTCTTCTAGGTGCTGAC--GCACCAAGAAAGTGTGCGGGCA 225757  
QY 1082 CGGCGAATTCCTGTGGCGGCTGATACCACTACCAATATGATGCTCTGATTCAGC 1141  
DB 225758 ATCCCTAGTAAAGTGAACGTGGGACCGGCGTCAATGMAACGCGCGCGATCG 225817  
QY 1142 AGACCGGCTTGTGACGCTGAACGAGATGCGGTGCTGAAGACTGGAAGCTGAAATATG 1201  
DB 225818 AGACCGAGCTTTCAGAGGACCGCGGAGGCAAGAGTACGTCTATTCGTCGATCC 225877  
QY 1202 ACGTGTGCGCACTTCTGTGGTGGGCGGACGTGCTGACGCGCACTGAAACCGGACA 1261  
DB 225878 TCGGCGGCAAAACGTGAGCGCGATGTCGTTCAATCCGCAAAACCGGATTCGCTACGCA 225937  
QY 1262 CCGGCAATTAATTCGTCGCGTGAACATGCTGCTAGATTAATGACGCTGATCAAG 1321  
DB 225938 ACAGCTGCGCTTGGCGGCAATGACAGGCGCAACCGGTACCTTCAAGAGGCGAAT 225997  
QY 1322 AGTTAGCGCTGACGCTTATTAACAACAGCGGACCGCAAACTGCGCGCGGCTTTG 1381  
DB 225998 GGTATCTGGGATGACCTTACCGATCCCTGGAAATTCGAGGACGCGCGGGGCGATC 226057  
QY 1382 AAAATATGGCGGATGACGCGATTAATACACACCGGCGGACCTTGTGTGCGCGG 1441  
DB 226058 TGAAG-----GGAATCGATCCGATGACCGGCAAGGCAAGTGGAGGGGCG 226102  
QY 1442 AGCGGCTTGCAGCAATTAATCGCGCTTTTGTGACAGGCGAGGTGTGTTCAACG 1501  
DB 226103 CGCGGACATCCCGGCTTTCGCGGAGTGTGACCGCGGGGCGGCGTGTGTTCAAG 226162  
QY 1502 GCGGAGCCGACCGCTATTTCCGTGCTTACGACGAAACCGGCGAATTTGTGGCAG 1561
```

Db 226163 GCGCCCTGACCGGCGAGTTCGAGCCCTTGATGCCGACACCGGCAAGAAGCTCTGGCACT 226222
 1562 CCCGCTTTGGACGCGTTCGCGACGCGGCGAGCGATCACTGACAGTTGACGCGCTGCAT 1621
 Db 226223 TCCAGACCCGCTCCGCGATCGAGGCAACCGGTGACGTGCGACGATGACGTGCAAT 226282
 1622 ATATGCGCANTC 1632
 Db 226283 ACAATCGCCGTC 226293

RESULT 8
 AP005957/c 299910 bp DNA linear BCT 12-SEP-2003
 LOCUS 23/31..
 DEFINITION Bradyrhizobium japonicum USDA 110 DNA, complete genome, section

ACCESSION AP005957
 VERSION AP005957.1
 KEYWORDS GI:27354278
 SOURCE Bradyrhizobium japonicum USDA 110
 ORGANISM Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE 1
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,
 Kohara, M., Matsumoto, M., Shimo, S., Tsurutoka, H., Wada, T., Yamada, M.
 and Tabata, S.
 TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110
 JOURNAL DNA Res. 9 (6), 189-197 (2002)
 MEDLINE 22484998
 PUBMED 12597275
 REFERENCE 2
 AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiyama, T.,
 Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K.,
 Kohara, M., Matsumoto, M., Shimo, S., Tsurutoka, H., Wada, T., Yamada, M.
 and Tabata, S.
 TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110 (supplement)
 JOURNAL DNA Res. 9 (6), 225-256 (2002)
 MEDLINE 22485002
 PUBMED 12597279
 REFERENCE 3
 AUTHORS Kaneko, T.
 TITLE Direct Submission
 JOURNAL (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
 Institute, The First Laboratory for Plant Gene Research; 2-6-7
 Kazusa-kamata, Kisarazu, Chiba 292-0812, Japan
 (E-mail: kaneko@kazusa.or.jp / rhizobase/
 URL: http://www.kazusa.or.jp/rhizobase/
 Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

FEATURES
 source 1..299910
 /organism="Bradyrhizobium japonicum USDA 110"
 /mol_type="genomic DNA"
 /strain="USDA110"
 /db_xref="taxon:224911"
 7..1380
 /gene="b1r6001"
 7..1380
 /gene="b1r6001"
 /note="ORF ID: b1r6001
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BACS1266.1"
 /db_xref="GI:27354279"
 /translation="MVDHPKRLDIPGAVMLGQYLARGVSVLVPMVEQAIVDPVRLGLD
 ALVVYARPNVLDLMSPAKAGLALVLDTEGVLAKEGNSPPMAALHVKSGYADI
 LGVFMGRHDAFLTAGTMEKELHLCGRPDPAARFALDGEPRGLVNAAN
 PVLNSRFGKCGGCEAMVRAGWDAAVYDRIADLKVFAVYLAIEDIRLAARPDRT
 ILVRPFESSEDEVYRALRAHNAVYIDGTGSLDRIKRAAAVYHLNCGIAVESVLGK

gene
 CDS
 7..1380
 /gene="b1r6001"
 /note="ORF ID: b1r6001
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BACS1266.1"
 /db_xref="GI:27354279"
 /translation="MVDHPKRLDIPGAVMLGQYLARGVSVLVPMVEQAIVDPVRLGLD
 ALVVYARPNVLDLMSPAKAGLALVLDTEGVLAKEGNSPPMAALHVKSGYADI
 LGVFMGRHDAFLTAGTMEKELHLCGRPDPAARFALDGEPRGLVNAAN
 PVLNSRFGKCGGCEAMVRAGWDAAVYDRIADLKVFAVYLAIEDIRLAARPDRT
 ILVRPFESSEDEVYRALRAHNAVYIDGTGSLDRIKRAAAVYHLNCGIAVESVLGK

gene
 CDS

gene
 CDS

gene
 CDS

gene
 CDS

gene

LPLQLEYLNTPTAGHALLPARVSRVNSFDELLGALDIHERETETEFDEPARVHAADIE
 APFHINDGOAERVADVLTDANSRRPVSLLATVGTGDKDSIGQIVGAASAVGS
 ANTERLRSRNPARRDKRIEPIVFRILLRLRIAHDAHSAQAQFATPAHCAATGLPLAS
 LAIEPAR"
 1384..2787
 /gene="b1r6002"
 1384..2787
 /gene="b1r6002"
 /note="ORF ID: b1r6002
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BACS1267.1"
 /db_xref="GI:27354280"
 /translation="MATSARTLITLLAAYQTFWIPVAQRRLRAAGHDELLAFDPS
 AEMSAAGVPTVMYREGKAGSPDRDAFDARVAYAGLDGNPLFSHERPTGFI
 TALRRPMTVYANAMETVLDLEGHKQAEIVLGGPISVIAFPYARRRNPIMFI
 EPSFRGKMTFTDTIGADPVMSPADVVSPEVRAVLDPTLQRAVIFKQCHYSA
 AFKRVNLRNARLTERLMDQFALGQESGHVLRARVYAAALNATLRKLKXIP
 DAPFVYVFFVADMALTLRSPDYDQVAIVDFLRTFIDSHVLLVKEHPAOTGAISA
 DRUFELARRPDNFPVLPQTNNTYLRADAVISVNSKGAELLLGKVVWGDVAFY
 RSCPLVYAVDLVDFARLREALSAGPDPDARGAPYFESAMWSYDELYISDPKLLD
 TPAASLRRAIAEERAVN"
 2799..3548
 /gene="b1r6003"
 2799..3548
 /gene="b1r6003"
 /note="ORF ID: b1r6003
 probable glycosyl transferase"
 /codon_start=1
 /transl_table=11
 /protein_id="BACS1268.1"
 /db_xref="GI:27354281"
 /translation="MPLVSTITPSNVEGLIERTISVOSQTFADWELLIADCSIDR
 TPEIVADIGERDPRVXLRKPNKGFPALRQAIDSAQRYLAFLOSDLMWPKLER
 QLAFASEKRALSTYAFRIINETNTVGLISVPALITVGLIKNTAIVLAADVRE
 IAGPIVKNKGVDYDFCWLCSILKRGHTVAGLDLARVYRGSVSSRRVRSKAWVQ
 IYVNEHLSTVRSWCPFAHMGARAWLKRRBP"
 3706..4668
 /gene="b1r6004"
 3706..4668
 /gene="b1r6004"
 /note="ORF ID: b1r6004"
 /codon_start=1
 /transl_table=11
 /product="rmp-glucose 4-6-dehydrogenase"
 /protein_id="BACS1269.1"
 /db_xref="GI:27354282"
 /translation="MPFESYKSRILVVGAGFISGICERLLDAGAEVVSADNYTG
 SRRIATLIANPLFEAVRHDPYPIEVDAIFNLCPASPIHYQSDPVQTTSTVHG
 AINMLGLAKRLKRIFOASTSEVGPPLIHPTEDYMGVNPPIGISCYDEGRCAET
 LFPDYMRHGLPIKVARIENTVGRMOPNDGRVSSFIYVALGGERITVFGDGQTRS
 FCYVDVYEAIMLMTTKEDIYGPINIGNSFETIELEKTELTGSRSKLVFKLP
 QDPRQRQPDLTAKTALNWEKVALDGLKETIAYFKSLEIA"
 4711..6024
 /gene="b1r6005"
 4711..6024
 /gene="b1r6005"
 /note="ORF ID: b1r6005
 unknown protein"
 /codon_start=1
 /transl_table=11
 /protein_id="BACS1270.1"
 /db_xref="GI:27354283"
 /translation="MSAPSGIMQHLQSRLLDSKPALTIILLHSVNTCLSIKVFQFS
 YIHSGRHWIAVAVVAPSVSLFPAARFSFGYAGAFYETMILGFLMIDVFSYS
 YDRLLAGLSAAASIVITLLPYFVRAPFQIVTLNVAAGEHLLTALIVSGTIAVAS
 TYNFRVLAINTIDYHDALFPCAAYILIGVSTSLIPPTACYSILGRMRAGVLI
 LLFVFLVITLTKFAFTPAVLTILAVLSRPLEARTSVITSEFPHLLGLVIVTGA
 INSIPGCFPLVILRMATSSALDYNNFPAHPLVTCQISMLKPVWPCAYODQUS
 VVNETYGFGEIMINASFATGEVAVSGVYLAFLALVAGETILAGNPAASGLPQRLVLI
 SSGTLVPLNTLPTVAMVTHGATLLFLIMVYVPRISIFORPGIIPD"
 complement (6021..7199)

Query Match 6.8%; Score 118.2; DB 1; Length 299910;
Best Local Similarity 49.4%; Pred. 6.1e-12;
Matches 433; Conservative 0; Mismatches 423; Indels 21; Gaps 4;

50 CCGCGCCCGCGCATTCGCGCAGGTACCCCGATTACCGCATGAACCTGCTGGCAACCCGC 109

Db	240677	CCGGTGC	CCGGCCCGCATCGAGATTATTTCTCCGGTCA	CCGGCAGCGCTCGAGATTCGGG	240618
Qy	110	CCGCTGTGAATG	ATTACTTACGGCCGCAACCAAGAAACTATTCGCACTCGCCCTGA	169	
Db	240617	AACCGAGCAACT	GTGATGTCTATCGGCCCACTTATGACGGGCAAGGCTTACAGCCCTCTG	240558	
Qy	170	CCGAGATCACTG	CCGACACGTTGATCAGTTGCACTGTGTGGGCGCGGAGATGAGG	229	
Db	240557	ACCAAGATCAAC	ACTTCGAAACGTAAGGGTCTCAGCCGCTGTGACCTTTTGCAACAGCG	240498	
Qy	230	CGGGGGCCGTAC	--AGGTACAGCCGATGATTCATGATGCGTGTATCTGGCAACC	286	
Db	240497	TCGTCAAGAGCC	ACAGAGGCGCGCCGATGTCACAAATGCGGTATGTTGTGGCGACCC	240438	
Qy	287	CCGCTATATGAT	TCAGAGCGCTGATGCGCAACAGCGCATCTGATCTGGGAACACCGCC	346	
Db	240437	CGATGGGAGGT	GTGATCGCGCTGAAACCGGAAACCGGCGACGATCTGGCGCTTACAGC	240378	
Qy	347	GCCCAATGCCCC	CGCTGCGCACGCTTAAACGCCCAAGCGACCCGACAGCGCGGCTGCGCC	406	
Db	240377	GCGAGCTTCCCA	CG-----ATCTGTTTCAGCTTGATTCGACACAGCGCGGCTCGGCT	240324	
Qy	407	TTTACGGCACGA	GCGCTCTATTTTCACTATCATGGGACAAACATCTGATGCGCTCGATATGG	466	
Db	240323	TGTGGGAGGACA	AGGTCTATCTCGCACACCGACGATCATGTCTGCGCTGACGCCA	240284	
Qy	467	AGACGGGACAGG	TCGATTCGATGTCGAACGTGATCGGGGCAAGACGGCTTACACAGTA	526	
Db	240263	AGACCGGCAAGT	TGTGTGTGGACACACCAAGTGTGAGACTACAGAAAGGCCAGATATTA	240204	
Qy	527	ACACACGGGGCC	GAATTGTGCGCAATGGCGTATCTGTGCGGGTTCCACTGCCAATATT	586	
Db	240203	CCCTGATGCGCT	GTGATGTCGCAACGGCAAGTCACTGTGTGCGGCTCCGGCGGC--GAGT	240147	
Qy	587	CGGCCATAGAT	AGCTTTATCTCGGGGACAGATTCGCGGAGCGGGTGAAGACTGTGGCGCA	646	
Db	240146	TGCGGCTGTGCG	GTATGTGTCGCTTATGACGCCAAGACGCGCAAGAGTGTGGCGCA	240087	
Qy	647	ACCACCTTTAT	TCCGAGACCGGGCGAAGAGGTGAAGACTTTGGGGCAATGATTTTCGAGG	706	
Db	240086	CTTACACCATTC	CCCGCGAAGGCCGAGCCCGGTTCACGACCTTGGCAGGGCGATGACTGGA	240027	
Qy	707	CGCGCTGATGA	CCGGCGTCTTGGGATCAGATACCTTATGATCTCCGTACGAACTTGTGT	766	
Db	240026	AGAACCGGCGCG	CGCTGTGCTGATGACCGGCAATTACGACAAAGACACCAAGACGATCT	239967	
Qy	767	TCTATGGCTCA	CGACCGGCGTGGGCGCCAGTCCGAACCCAGCGCGGACCGCGCGGCA	826	
Db	239966	ATTGG-----	GGCTGTGGCAACCGGGCGCGGTGGCGGGGAGAACCCATCTCCGGCG	239916	
Qy	827	CGCTGATGAGCA	CAACAACCGCGCTTTCGCTGCGTCCGACACGGGCGAGATTGTCTGGC	886	
Db	239915	ACATCTTACAC	CTCTCGTGTGCTCGCGTCTGATCTCGAACAACGGCGAATCAAGACTT	239856	
Qy	887	GTCACGAGACCT	TGCGCGGCGGACCAACTGTGGGACCAAGA	923	
Db	239855	ATCACCGATACA	CCAGAACGACTCTTGGACTGGGA	239819	
RESULT 9					
AP005936/c					
LOCUS					
DEFINITION					
AP005936					
Bradyrhizobium japonicum USDA 110 DNA, complete genome, section					
231.					
ACCESSION					
AP005936 BA000040					
VERSION					
AP005936.1 GI:27348543					
KEYWORDS					
SOURCE					
ORGANISM					
Bradyrhizobium japonicum USDA 110					
Bradyrhizobium japonicum USDA 110					
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;					
Bradyrhizobiaceae; Bradyrhizobium.					
REFERENCE					
1					

AUTHORS Kaneke,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Matanabe,A., Igesawa,K., Iriyuchi,M., Kawashima,K., Kohara,M., Matsumoto,M., Shimo,S., Teunoka,H., Wada,T., Yamada,M. and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110

JOURNAL DNA Res. 9 (6), 189-197 (2002)

MEDLINE 22484998

PUBMED 12597275

REFERENCE 2

AUTHORS Kaneke,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiumi,T., Sasamoto,S., Matanabe,A., Igesawa,K., Iriyuchi,M., Kawashima,K., Kohara,M., Matsumoto,M., Shimo,S., Teunoka,H., Wada,T., Yamada,M. and Tabata,S.

TITLE Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110 (supplement)

JOURNAL DNA Res. 9 (6), 225-256 (2002)

MEDLINE 22485002

PUBMED 12597279

REFERENCE 3 (bases 1 to 300600)

AUTHORS Kaneke,T.

TITLE Direct Submission

JOURNAL Submitted (20-NOV-2002) Takakazu Kaneke, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 222-0612, Japan

URL: <http://www.kazusa.or.jp/rhizobase/>,
E-mail: kaneke@kazusa.or.jp
URL: <http://www.kazusa.or.jp/> (ex.2318), Fax: 81-438-52-3934

FEATURES

source 1. 300600

1. /organism="Bradyrhizobium japonicum USDA 110"

2. /mol_type="genomic DNA"

3. /scrn1="USDA110"

4. /db_xref="taxon:224911"

5. /complement(28..309)

6. /gene="bsl0296"

7. /complement(28..309)

8. /gene="bgl0296"

9. /note="ORF ID:bgl0296"

10. /unknown_protein"

11. /codon_start=1

12. /transl_table=11

13. /protein_id="BAC45561.1"

14. /db_xref="GI:27348544"

15. /translation="MRLIPVGVVACVALSGTASARGDGSHMSAGALGTSAAAPGNSLGLPSSDSGRATKGPLGTDPIDDDARLEKMLEGSLCRGC"

16. 1017..2390

17. /gene="bly0297"

18. /complement(28..309)

19. /gene="bly0297"

20. /note="ORF ID:bly0297"

21. /hypothetical_protein"

22. /codon_start=1

23. /transl_table=11

24. /protein_id="BAC45562.1"

25. /db_xref="GI:27348545"

26. /translation="MAQRPPYASLPGRGRTRRRVYVNPAPAPMRRLITGRGTYSISGNAMRLGKLVISCAVLAVGSASLEPYDPSADTEIRGINIMPYSGPASAYIGKMSAYERYINNGINGKXINPISYDVASSKXAEORARLVEDDEVIALPAPGTA SNAALOKVYNNRHVPOLPVITGASHRADDEHPMTIGMLPSRTETARVIALYILHHHPGARIGLIHQNDPGRDYLLGLKDVLQDKYAKQVVASLPIEVESEPTIDSQVVAISKAPDIFIDATKFAAQAIRKLAEDWPHIHIVNVSQVGLVKEPAGLIDNAKGLSKAQEMVTQVQADSHQMHFRAPAKYPEADRSSEGPLVAFNMSALVEVIRCGDGLTH ERIMVQVADLDELLEAVIPGIRIKITSPDTFHDIEOLQMRFTGESMEWFGPLDGRAD QATPLPD"

27. 2353..2988

28. /gene="bly0298"

29. /complement(28..309)

30. /gene="bly0298"

31. /note="ORF ID:bly0298"

32. /unknown_protein"

33. /codon_start=1

34. /transl_table=11

gene /protein_id="BAC45563.1"

CDS /db_xref="GI:27348546"

1. /translation="MDALTRRRPCPTRTSLMDPESMEVAPRRLAKTKMVFVSTLRIRAEHVSRAQGRVRRFLKMFDDTEALVGRIRGRSHDCLDPSRORAEVQCGICUSASRGILQARLRPCSRRCVRNRLPAAAMRAHIGLISRTSSPPVADMK LIVSTFRILTSRKEGRRLTRPPRLPAPHVSRTHRCSGALTTIGEVSK"

2. 3261..6173

3. /gene="bly0299"

4. /complement(6919..8190)

5. /gene="ragd"

6. /complement(6919..8190)

7. /gene="ragd"

8. /note="ORF ID:bly0300"

9. /codon_start=1

10. /transl_table=11

11. /product="RagD protein"

12. /protein_id="BAC45565.1"

13. /db_xref="GI:27348548"

14. /translation="MSDVERESDDDEADRPGEASRLRYELPGKVARSKRRYGGPLLG GYVILLAGLIGIGRWNYEAREBAVAEVOFTNPEPVAIVQASGLMKVTLAT TTAPEANIPARTSGYIEKRYVDIGRVKKGDLLVETAPEDQQIAQAQATLAAQQA ALQQAASREPLADVTNSRDSNTVYKQGMILAAQGDNDRLTLPAQQAAGVAAQSLAQA AQIRVLAQEKSYORVVAAPPDGYTQGNUNGSLYQASGTFMTMMSNVTIRQVFPDQ DEAPGVAAGVDSIDIRYELPGRTPAGVTRATLALPGSRITLITEIDVNPDSGLSPG IYCTVELSLIPRRPTSTIPSDALVFDQNGLHVAVNGRVTHFOQVLSRDFGTVFVR DGVPQGVVLANFAVNLAEQSKYTVKSEVS"

15. 1017..2390

16. /gene="ragC"

17. /complement(8187..11369)

18. /gene="ragC"

19. /note="ORF ID:bly0301"

20. /codon_start=1

21. /transl_table=11

22. /product="cation efflux protein"

23. /protein_id="BAC45566.1"

24. /db_xref="GI:27348549"

25. /translation="MGIYRPAALDPHTVYVLAALIFLIGIAIRSPMDIFPPIRPV LVYIWSYTGSLSTEMEORVSTYSQISANVTGIRKITEQTLNGLSVQKIRYQDPNVL DLAISQVSAVNAIRALMPGLOPPIIVQFNASSVPLQLSLESNLSNCOQLDFEY RVRQOALAPVGVTLPTPAGKRYQIMWDIDPNKLSRGLTPDIYVNAVYQNTLPTG TTKIGDQVYRTNATPATIODLNMIPVKFANGATFLKDVAQVDRGAQVOQVIVED GHRVAIVSYKNGNASTLAIVNGVKALISIRASAPAGKINELFQOSIFVHSVNGV LRGCATLACGLTALMILVPLGSWSTLVYLSIPLMSSLVLYLFGELTNMTTGLGL ALAVGILVDDSTYTITENTHRLWTBEGMPLSEATLHGAATAPTYSTALAGVPTSV VLEEGPARKYLPVGLAVVPAMLASTGSLSTLPTITGLLKGRRHGGGRSNTFA RASAAFERGFEHLRDGYSDLLTLARVIVPVAVLVALGATMVFVYGRDPLID GGOIOLHVRAPAGTRIERTEALFOAVEDKIREVIRPDBALVDNIGLARPYNLAF T DGSITIGVNDGTLIVSLKDGKHPADYIRKLRLQVLPAPFEDTFYFOAADIVQIILNFG

LPAQIDVITGVYGSNNLAVAKELQSLAIPGVDAHLQOEVDAPAFADIDRTSAQ
LGIATSTAVINIVSLSSAOVSPNMTDPTSGIPYILAVOTPEPRLANSIGMTPV
LSLAAGTGVPELISNVAPFKVGTPTVNOANVPYDVVAASQDQIGSVADID
KUTSTIOKLOLPENSIOVAGQIOSHOSRDIGIGLLPAIIVILMTYNTYNDPDE
VIMALPATFCGIVLTVFTGTTLVNPSLNGAIMAVGANSILVIFARBOUKGH
SAFOALSAGRTIRIPVMTAAIVGMIPMAIGGPEBONALARAIVIGLLFATPT
TLIIVYLFPAMLRKNDGKPRHGFEEPOE"
/gene="rpoH3"
/complement (11449, .12372)
CDS
/gene="rpoH3"
/complement (12433, .13857)
/note="ORF ID: b110302"
/codon_start=1
/transl_table=11
/product="RNA polymerase subunit c"
/protein_id="BAC45567.1"
/db_xref="GI:27348550"
/translation="NOTSHVARSASVAAAGAAVAPFISAAIRYELLPEQEQ
QLARRHETRDGADALVTLRLAKLARGKGLPMVLIAEALGLVTAARF
EPGRGARFSTYVIMIKAAIHEVILNSGLVIGITTAQKLFELREIRKATSDV
SGITPDVAELIAGKLDVLTAREVILEMDARLNGMSINARKESEGTLEALIVDGV
AETMLADHEOTERAKALRVALLGLAARHVFEARLTCEPVLIDOLARELISSE
VQOETIRAPAKYRAVMLAADAAPRAAVCNV"
/complement (12433, .13857)
gene
/gene="ragB"
/complement (12433, .13857)
CDS
/gene="ragB"
/note="ORF ID: b110303"
/codon_start=1
/transl_table=11
/product="two-component sensor histidine kinase"
/protein_id="BAC45568.1"
/db_xref="GI:27348551"
/translation="MHQSQLIRSSITFLMALAAVAFVVGAFARIWKLDYLVAR
SRMTITRIHFTADLPARRISALADHLQDSRGVYAGLPAAGTRLAGNIDVPRE
LGLDVGAVGRLDLEKPAHAYVVAVRGLDGDVLMGNVDETEISIVGQAL
ALGILPAFLCTILAGMALSIRAPRVEENRORVILVGELEPRLPEGDTPFRLA
GVNGLDMEHTMINALAGVNDIADHATPITRVALERETRAATLEDDOITDK
AAGIDQSLATVIALRLTEIKNRMAFGFVALDEILEVCDVETPAEDKRIALG
VIERNVQWGRDLDLEAVANLVDAVAFPTSGRVELALEDEGTALVRSDIPG
LREOEVALRFRYSDKRNTPGVGLGLVAAIVKLHGFRLIVGAPAPGREITLAW
TRADKAATCRRTLHADADLMG"
/complement (13867, .14841)
gene
Query Match 6.3%; Score 110.2; DB 1; Length 306500;
Best Local Similarity 46.7%; Pred. No. 1.8e-10;
Matches 736; Conservative 0; Mismatches 763; Indels 78; Gaps 9;
QY CTGCTGCGAACCAGCCGCTGGTGAATGATTAACGACGCGCAACCAAAACTAT 153
DB 50383 CTAAAGCCGCAATCCGACTGAGCCAGTGGCCGCTCCGCGCAACGACTACGCGCAACAG 50384
QY 154 CGGCACTCGCCCTTGACCCAGATCACTGCCGCAACGTTGTGTCAGTTGCACTGTGCTGG 213
DB 50323 CGCTTCAGCCCGCTGAAGCAATCAACCGAAGACCGCGCAAGCTGACGCTCCCTTAT 50264
QY 214 GCCCGCGGATGAGAGCG---GGGCGCGTACAGGTCAAGCGCGCATGATTCATAGGCGTG 270
DB 50263 TCGTTTCGCTCGCTCGCTCGCTTGAACGAATCTCGCCGATCGTCTCGGTAAACAG 50204
QY 271 ATGTATCTGCGCAACCCCGGTGATGTA-----TCAGAGCGCTGGATGCGCAACAGCG 324
DB 50203 CTCTACGTTTCGAGCTCTCGGGCCGGAATACGTCTACGCGCTGATCCCGCCACCGGC 50144
QY 325 GATCATCTGCGAACAACCGCGCCCAACGCGCCGCTGCCACAGCTAAAGCCCAAGCG 384
DB 50143 GTGCGCAAAATGACCTACGAACCCGATATTCGGAAGAGTGTGTGAAATACCTGCTGCTGC 50084
QY 385 GACCG---CAAGCGCGCGTGCCTTTACGCGACAGAGCTCTATTTCAGCTCATGAGAC 441
DB 50083 GACGTAAACATGCGGCGGTCTCTATGCGACGCGAAGATCTTGTGTGCGCGCGCTCGAC 50024
QY 442 AACCATCTGATGCGCGTGAATAGAGACGGGCGAAGTGTGATTCGATGCAACGTGA 501

DB 50023 GCGAAGCTGAGCGGCGCTGATGCCGCGACCGGCAAGG-CGCTCTGACCGCGGAAGCTGT 49965
QY 502 TCGGCGCAAGACGCGCTTGAACGATTAACACACGCGGCGGATTTGTCGCAATGGCTCATC 561
DB 49964 CGATTACAGAGGAGT---CGGTCACTACCTCGCGCGCTCGTCTGCGCGACAGAGTC 49907
QY 562 GTGCGGGTTCACCTGCGCAATATTCGCGCTATGAGATGCTTATCTCGGGGACAGATTCC 621
DB 49906 ATCAACGGTTTCGCGCGCGGCGAGTACGCGGTGCGCGCTCGCTCCAGGCTTGAATC 49847
QY 622 GCGACGGGTGAGAGCTGTGCGCAACCACTTATCCGACGCGCGCGAAGAGGTTGAC 681
DB 49846 AACACCGCGAAGCAGCTGTGCGCAACCTTACACGCTTCGTCGCGGATGAGCCCGCGGC 49787
QY 682 GAGACTTGGGCGCAATGATTTGAGAGCGCGCTGATGACCGCGCTGCGGGGTCAATCAAC 741
DB 49786 GACACTGGAAGGCGACTCGCGCCAGCATGACGCGCGGTGCGCTGCTGTGCTGCTCC 49727
QY 742 TATGATCCCGTGAAGAACTTGTGTTCTATGCTCGACCGCGGTGCGCGCAAGCTCCGAA 801
DB 49726 TACGATCGAAGACCGACACGCTCTATTGGGGTACGACATCCCGGTCCGTGGAACAG 49667
QY 802 ACCCAGCGCGACCGCGGCGGACGCTGTATGGACCAACACCGCT----- 850
DB 49666 GCGGTACGCTGACCGGCGGACGCGCAATTTGCGCAAGCTGACCAACCTTACACCGCTCG 49607
QY 851 ----TTGCGGTGCTGCCGACACGCGGAGATTTGTGCGGTACGACGCTTCCGCGC 906
DB 49606 ACGGTGCGTTGATCCCAACACCGGCAAGATCAAGTGGCAATCCAGACCAAGCCCGGC 49547
QY 907 GACAACTGGAACCAAGATGACGTTGAGATGATGTCGCAACGTCGATGTGCAACC 966
DB 49546 GATGCTGCGGCTGTGACGCGGTCAAGAGCCGTGCTGCGGATTCGAAGATGCGCGGC 49487
QY 967 TCGGCGGAGATGAGAGGTCGCGCGCATCAACCCAGTGGGACGCGGAGCGCGCT 1026
DB 49486 AGCACCGTCCGCGCGCTGAT-----GAAAGCGAATCCGCAACGCGC 49448
QY 1027 GTGCTGACGGGTGCGCTTGCACAGCGGACGATGTGTCGTTTGAATGCGGCTCGGCG 1086
DB 49447 TACTTCTTCGCGCAATCGGAGACCGGCAAGGTGCTTTCGCGCGAATACTGTTC 49388
QY 1087 GAATTCCTGCGGCGCGGTATACCACTTACACCAATATGATCGCTGATCGACAGAC 1146
DB 49387 -----TCCAACTGGGCGCAAGAGTGGGACATCGCCAGATGCGC 49349
QY 1147 GCGCTTGAACGATGACGAGATGCGGTGCTGAAAGCTGACGTTGAATATGACGTC 1206
DB 49348 GCGGTGAGGATTCGCGACAGCGTCCGCGCGCAACATCCCGC-----AGGACATC 49295
QY 1207 TGCCCGACCTTCTCGGTGCGCGGACGTGTCGACCGCATGAAACCGGACCGGCGC 1266
DB 49294 TGCCCGAACCTGATCGCGGAGAACCTGCGACGCGATGTCTTCAACCGCGACCGCGC 49235
QY 1267 ATTACTTCTTGCAGCAATGCTGTGATGATATTAATGAGATTT 1326
DB 49234 CTGCTTACATCCGTCGCAATGTCTGATGACGTGCTGCTCGACGCTGCTAC 49175
QY 1327 AGCGCGCTGACGCTTAAACACGCGGACCGCAAACTCGCGCGGCTTTGAAAT 1386
DB 49174 AAGCGCGGTGTCTATCTCGCGCGCAATTCGCGACCAAGAGACCCCGCGGTTTC 49115
QY 1387 ATGGCGCGATTCGACGAGATGATATGACACGCGGCGCACTTGTGCGCGGAGCGC 1446
DB 49114 CTGCGCAGCTCTGCGCTTGGATCCGTCGCGAAGAGTCTGTGATCAAGAA 49055
QY 1447 CTTGCGGCACTTACCTGCGCGTTTGTGACAGCGACGCGGTGTGTTCAACGCGGCG 1506
DB 49054 GACCTGCGTTCAACGCGCGACGCTGACCAACGCGCGGTGCGCTGTTTGC CGGCAAC 48995
QY 1507 ACCGACCGCTATTTTCGCTCGCTTACGCAAGAAACCGCGCAAGCTTGTGCGAGCGCGCT 1566

Db 48994 ATCCATGCGGATTTCCGCCCATTCATGCGAAGTCCGCCAAGGTGTTGACGACGAAAT 48935
Qy 1567 CTGCGACGCGTCCGACGCGGACGCGATCACTAGAGTTGACGCGCGTGCATATATC 1626
Db 48934 CTGCGCTCCGCGATCCGCCAGATCCGATCTACACGATGACGCGCAAGCAATATGTC 48875
Qy 1627 GCCATGCGTCCGCGCGG 1643
Db 48874 GCCATGCGTCCGCGCGG 48858

RESULT 10
AF277373 2097 bp DNA linear BCT 07-MAR-2001
LOCUS Ralstonia eutropha tetrahydrofuryl alcohol dehydrogenase gene,
DEFINITION complete cds.
ACCESSION AF277373
VERSION AF277373.1 GI:9255864
KEYWORDS
SOURCE Waurensia eutropha
ORGANISM Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Burkholderiaceae; Waurensia.
REFERENCE 1 (bases 1 to 2097)
AUTHORS Zarrt,G., Schraeder,T. and Andreesen,J.R.
TITLE Catalytic and molecular properties of the quinohemoprotein
tetrahydrofuryl alcohol dehydrogenase from Ralstonia eutropha
strain Bo
JOURNAL J. Bacteriol. 183 (6), 1954-1960 (2001)
MEDLINE 2112557
PUBMED 11222593
REFERENCE 2 (bases 1 to 2097)
AUTHORS Zarrt,G., Schraeder,T. and Andreesen,J.
TITLE Direct Substitution
JOURNAL Submitted (13-JUN-2000) Microbiology, University of Halle,
Kurt-Mothes-Str. 3, Halle 06120, Germany
LOCATION/Qualifiers
1..2097
/organism="Waurensia eutropha"
/mol_type="genomic DNA"
/db_xref="taxon:510"
1..2097
/codon_start=1
/transl_table=11
/product="tetrahydrofuryl alcohol dehydrogenase"
/protein_id="AAF8635.1"
/db_xref="GI:9255865"
/translation="MHSNKLRLVLCVMAAASVALPAPMAFGANAARVDAALRAMEA
GTEPMSVGLDYAEFRFSKLEOVNAGNVTGLAWSYDLSEIRGVEATPLVVDGMYV
SAFMSVVAHLDARTGKRLMTYTPQVPRDAVYCCDVNRGVALVCGAFVGAEPGRV
VAIDATGKRWEDDTVDKSKSYITGAPRYNGKVIINGGAEYGVRTTAYDAE
TGKQQRWYTVPGDPAPRPENBMAAATAVDPGSKYVINGGGVNTVMAFDEPLNL
MYITGNAGPMSKSLSPKSGDNLVAASVALNPDTGEVVMYQTPGDNNDYSTOD
IILADKIDGQPRKVIILHAKNGFEVIDRTNGKFIISAKNPVDMAGYKNGPVE
TPQADTSGKADVPGPFGAHMWSHFHKLGLAFIPQAHVPLILDNKEMVHQKD
SPFAHSGVMNLGMLVNAAPRPSKMGRIILMDPLAOKAVMRBDIAGMNGGLTAG
NLFVQCTADGRVLAHAAATGKELMQAPFTSSGVAAPVYILDRGYVVAAGMGVYG
LSGRGSDQAPGKIVTVLNGKTPMPASAPYAAVLPUSGVKIDPAKVDGQKLYVNC
VFQGVGVADKGNINFLAYVGAPMIEHLKILFNGPVGKMPDTKLTAEDEVKI
QAFICGTADAVPRKP"

ORIGIN
Query Match 63%; Score 110; DB 1; Length 2097;
Best Local Similarity 48.8%; Pred. No. 3,4e-10;
Matches 450; Conservative 0; Mismatches 430; Indels 42; Gaps 4;

Qy 119 AATGATTACTACGCGCCGCAACCAAAAATATGCGCACTGCCCTGACCCAGATCA 178
Db 143 ACTGGCCACGCTACGCGCTTGATTAACCGAGACCGGCTTCAGCAAGCTCGACAGTCA 202
Qy 179 CTGCGCAACAGCTTGTGCTGCACTGTCTGCGGCCCGCGGAGTGA---GGGGGGG 235
Db 203 ACCGGGCAACGTCAGGAACCTGGGGCTGGCGTGTCTTACGACCTGGAATGACGCGCG 262

Qy 236 CCTTACAGGTCAAGCCGATGATTCATGATGAGCGTGAATGATCTGGAACCCCGGATG 295
Db 263 GCGTGAAGGCGACGCGCTGTGTGTGTCACGCGCTGATGATGCTTCGCGCGTGAAGG 322
Qy 296 TGATCCAGGCGCTGATGACGCAAAACAGCGCATCTGATCTGGAACAACCGCGCAACTGC 355
Db 323 TGTGTCATGTCATGACGACGCCGACCGGAGAGCGCTGTGGAACCTACGACCCGCAAGTGC 382
Qy 356 CCGCGCTGCGCAACGCTAAAGCCCAAGGCGAC---CGCAAGCGCGCGCTTGCCTTTAG 412
Db 383 CGCGCGCAACAGCGCTTACAAAGGCGCTGTGTGCACTGTGTCATGCGCGTGTGTATCC 442
Qy 413 GCACGACCTCTATTTCAGCTCATGGAACAACCATCTGATTCGCGCTGATATGAAACCG 472
Db 443 AGGCGAAGTCTTCTGTGCGCGCTTGAACGCGCGGTGTGTGTCATGACGACGACCG 502
Qy 473 GCCAGTCTGATTTCGATGATGTCGAACTGTGATCGGAGCAAGCGGCTTGACCACTAACCA 532
Db 503 GCAAGAAAGTCTGGGAGCAAGAACATCGTCGACCGCAAGTCTTACACCATCAACG 562
Qy 533 CGGGCGGATGTCGCAATGCGCATGCGTATCGCGGTTCCACTGCTCAATATTCGCCCT 592
Db 563 GCGCGCGCGCTTCTACACGCAAGTATCATCGCAACGCGCGCGGATATAGGG 622
Qy 593 ATGATGCTTATCTCTCGGCGACGATTCGCGACGAGTGAAGAGCTGTGCGCAACACT 652
Db 623 TGGCGGCTATATCACCGCTTACGATCGCGACCGGCAAGGCAATGCGCTGTGTACA 682
Qy 653 TTATCCCGGACCGCGGCAAGAGGTGAGAGCTTGAGGCA----- 694
Db 683 CGGTGCGGCGGACCGCGCAAGGCAATTCGAGAACGACATGCGCAAGGCGCGGCA 742
Qy 695 -----ATGATTCGAGCGCGCTGATGACCGGCGCTGTGGGCTGAG-----A 736
Db 743 CTTGGATTCGACGAGGCAATGATGATCAACGCGCGCGGCAAGTGTGAACACA 802
Qy 737 TCACCTATGATCCCGTACGACACTTGTGTCTATAGCTGTGACCGCGCTGTGGCGGCGCT 796
Db 803 TGGGTTGATGATCCGACCTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 862
Qy 797 CCGAAACCCAGCGCGGACCGCGGCGGACGCTGTATGACCAACACCGCTTTGCGG 856
Db 863 GCGCGACGCTCGCGACGCGCGCAAGGATGACCAACCTTACGCGGCTCGGTGTGGCGG 922
Qy 857 TGGCTCCGACAGCGGCGGATGTTGTGCGGTACACGACCTTCCGCGCGGACAACTGGG 916
Db 923 TGAACCCCGACACCGGGAATGATGCTGTGCACTACGAGAAACCCCGGCGGACACTGGG 982
Qy 917 ACCAAGATGACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 976
Db 983 ACTACCTGACACCGACGACATCATCTGCGACCTGAAGATGACGCGCGCGCGCA 1042
Qy 977 TGGAGGCTTGGCGCGCATCA 998
Db 1043 AGTGATCTGATGCGCCCA 1064

RESULT 11
LOCUS CTGHEBH 2959 bp DNA linear BCT 12-APR-1996
DEFINITION C.testosteroni qheh gene.
ACCESSION X81880
VERSION X81880.1 GI:663195
KEYWORDS qheh gene; Quinohemoprotein ethanol dehydrogenase.
SOURCE Comamonas testosteroni
ORGANISM Comamonas testosteroni; Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
REFERENCE 1 (bases 1 to 2959)
AUTHORS Stoorvogel,J., Kraayveld,D.E., Van Sluis,C.A., Jongejan,J.A., De
vries,S. and Duine,J.A.
TITLE Characterization of the gene encoding quinohemoprotein ethanol

dehydrogenase of *Comamonas testosteroni*
 Eur. J. Biochem. 235 (3), 690-698 (1996)
 MEDLINE
 PUBMED
 8654419
 2 (bases 1 to 2959)
 Stoorvogel, J.
 Direct Submission
 Submitted (16-JAN-1995) J. Stoorvogel, Department of Microbiology
 and Enzymology, Delft University of Technology, Julianalaan 67,
 2628 BC Delft, NETHERLANDS

FEATURES
 source
 1. .2959
 /organism="Comamonas testosteroni"
 /mol_type="genomic DNA"
 /strain="ATCC 15667"
 /db_xref="taxon:285"
 /cell_line="E.coli JM109"
 350. .354
 361. .2487
 /gene="qheh"
 361. .2487
 /gene="qheh"
 /codon_start=1
 /transl_table=11
 /product="quinohemoprotein ethanol dehydrogenase"
 /protein_id="CA57464.1"
 /db_xref="GI:661196"
 /db_xref="GOA:Q46444"
 /db_xref="Swiss-Prot:Q46444"
 /translation="MRLLDNSHGFGRNVTLLACGSAAPFATGPAQAQAAVOR
 VGGDFIRANAARTPWPITIGVDYAEIRYSRLQINAAVKDGLAMSVNLESTRGVEA
 TPVVDVIGIVVSAWSVVAHIDRTGNRIWTDPOIDRSTGKGCDDVNRGVNLKGA
 KYVVGADRLILADATGEVWHONTFEGQKSLITIGAEPVFKKXVLIIGKGAIEG
 VGGYIADYAEIERKEMFVGDSPKFEDESKRAATDPGKMMWAGCGGTMM
 DSWTPAEINTWVGTNGSPMSHKVRSFGGDNLYLISIVLADPDQIKYKWHVETP
 GKNWDYTSQPMILADIKTAGKPRKRIILHAPKGFPTVIDRNGKFIKRNVPYNNMA
 SGTDRKRPITGLAADGSKPDVAVPGTGANMHWMSFNPTGLVYLPQAVPNVMA
 DDKRMFNQAGPKQSGTGMNTAFENAPKSPKPFGLLMDVPAKAAVSHVS
 PNVGTLITPAGNVFQGTADGRILVAHATGEKLEWAPRTGVVAAPSTYWDGQYV
 SVAVGWGVGLAAERERQGGTGYTVTVVAGKAMPFVAGRTQLLQGVYVDAKY
 EAGTMYLVANVCFCHGPGVDRGNTPNLVYDASYLENLPFVFKGPAWVGMDPT
 GLSGDDVESLKAFLQGTADAIRPKP"

sig_peptide
 361. .453
 /gene="qheh"
 454. .2484
 /gene="qheh"
 /product="unnamed"

ORIGIN

Query Match 5.8%; Score 101.2; DB 1; Length 2959;
 Best Local Similarity 47.5%; Pred. No. 1.3e-08;
 Matches 439; Conservative 0; Mismatches 443; Indels 42; Gaps 3;
 117 TGAATGATTAATACGCGCGCAACCAAGAAAATCTATCGCACTGCGCCCTGACCCAGAT 176
 534 TGAATGATTAATACGCGCGCAACCAAGAAAATCTATCGCACTGCGCCCTGACCCAGAT 593
 177 CACTGCCGCAACAGTGTGTCAGTTGCACTGTCTGGGCCCGCGGAGTGAAG--GCCGG 233
 594 CATATGCCGCAACAGTGTGTCAGTTGCACTGTCTGGGCCCGCGGAGTGAAG--GCCGG 233
 234 GGCCTGACAGGTGACGCCGATGATCATGATGCGCGATGATCTGGCAAAACCCGGGGA 293
 654 CGCGCTGAGAGCCACCGCTGTCTGTGTGACGCGATCATGTATGTAGCGGCTCATGAG 713
 294 TGTATCCAGCGCGCTGATGCGCAACAGCGCATCTGATCTGGGAAACACCGCGCCACT 353
 714 CGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 773
 354 GCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 410
 774 CGACCGAGACCGGCTTCAAGGGGTGTGCGACGTGTCAACCGCGGCGTGGCGTGTG 833

411 CGGACGACGCTCTATTTCAGCTCATGAGCAACACATCTATGCGGCTGGATATGAGAC 470
 834 GAAAGGCAAGGTCTATGTGGGGGCGTGGAGATGGCGGCTTATGCGCTGATGCGCCAC 893
 471 GGGCAGGTCTGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 530
 894 CGGCAAGAGGTCTGAGCAACCAATATCTTCAAGGGGCAAGAGGGGTGTGCTACATAC 953
 531 CACGGGCGGATTTTCCGCAATGCGCTCATGCTGTGGCGGTTCCACTTCCCAATTGGCC 590
 954 CGGCGCGCGCGCTGTTCAGGGCAAGATCATGATGCAAGACCGCGCGCGCAATTATG 1013
 591 CTATGATGATCTTATCTGCGGGGCAAGATTCGCGCAAGGATGATGATGATGATGATGAT 646
 1014 CGTGGCGGCTTATCTACCTGCTTATGACCGCGAAGCCGGGAGGGAAATGGCGCTGTT 1073
 647 -----ACCACTTATCCGCAAGCGCGGCAAGA 674
 1074 CAGCGTACCCGCGGATCCCTCCAAACCTTCAGAGACGATCATGAAAGCGCGCGCCAG 1133
 675 GGGTGAAGAGCTTGGGGCAATGATTTGAGGGGCGCTGATGACCGCGCTGCGGCTCA 734
 1134 GACCTGGACCCCGCGCAATATGTGGAGGCGCGCGCGCGCACATGTGGACAG 1193
 735 GATCACCTATGATCCGTGACGAACCTTGTGTTCTATGCTGACCGCGCTGGGCCAGC 794
 1194 CATGACCTTGCATCCGAACCTCAACACCATGATGCTGGGACCGGCAATGATGCTGCTG 1253
 795 GTCCGAACCGAGCGCGGCAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 854
 1254 GTCCGACAGGTTGGGCGAGCGCCCAAGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1313
 855 GATCGTCCGCGACGCGGCGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 914
 1314 CCTGATCCGACACCGGCAATATCAAGTGGCAATATCAAGAAACACCGGCGCAACTG 1373
 915 GCACCAAGATGACGCTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 974
 1374 GGAATTAACCTCCACCGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1433
 975 GATGAGAGGTGCGCGGCAATCA 998
 1434 CAGGTCTACTGATGCGCCCA 1457

RESULT 12
 AF176640
 LOCUS
 DEFINITION
 Pseudomonas stutzeri sigma54-dependent transcriptional regulator
 AdhR (adhR), alcohol dehydrogenase (adh), coenzyme PQQ synthesis
 protein A (pqdA), and coenzyme PQQ synthesis protein B (pqdB)
 genes, complete cds.
 ACCESSION
 AF176640
 AF176640.1 GI:9957180
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas stutzeri
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE
 1 (bases 1 to 7097)
 Chang, C.-H., Herrick, J.B., Okinaka, R.T., Brainard, J.B. and
 Terwilliger, T.C.
 Identification and characterization of genes activated by
 2-chloroethanol in *Pseudomonas stutzeri* BC-2
 Unpublished
 2 (bases 1 to 7097)
 Chang, C.-H., Herrick, J.B., Okinaka, R.T., Brainard, J.B. and
 Terwilliger, T.C.
 Direct Submission
 TITLE
 JOURNAL
 LABORATORY, MS-M888, Los Alamos, NM 87545, USA
 LOCATION/Qualifiers
 1. .7097

```
/organism="Pseudomonas stutzeri"
/mol_type="genomic DNA"
/strain="BC-2"
/db_xref="taxon:316"
/complement(123..2027)
/gene="adhR"
/complement(123..2027)
/gene="adhR"
/codon_start=1
/translation="sigma54-dependent transcriptional regulator
  adhR"
/protein_id="AA09248.1"
/db_xref="GI:9957181"
/translation="MRRSDPAATPTSAKHIVDSOVIARSMORSNNGVGRDKLLFE
  APLVASGRKQIAERDILLDCAMPEQIQAALAPODWTLACDIEGVLSLRGAG
  TRLSOFVIRIGALQSEONVCTTAPGALAEKRSVFVGEHETLEAPRPSCLATPIE
  LDGIVLVDASSRQGSAPVYTESLLEPAKRIENRPAQJGEVILVLEHRLDVG
  PLEGLAPDPACKLRCAIRIAQQLHLEVRAGLDPAELFDMPEAHACYRRHDSAP
  LRLSHGGEHLACITQGPORFAPLAEPTPRRTASLPAHSGPCLQDPALQDAL
  AARAKAFADIPVLINGETGKEVFAARLHEGSAAGSAPVAINCSSIPALISEL
  PCHVEGAFTRSGRKGALGKLEQAHKGTLEIDEMPELOGRILRVIOENRSLG
  NOLIKYDCAITCAHRDLQOLSAEGFRPDIYLRINGLRVSLPALRESDIGALIOH
  LHEITGQOQLSPALQALDHYHMPKTRQLYOCLEAALADDEIYGHDPDE
  VVAGCGEQATVAGEGTLEAGSTIQLEAPRAIRAEVAGNISAAARSIGIARATL
  VKLKGMEEG"
/gene="adh"
/translation="alcohol dehydrogenase"
/protein_id="AA09249.1"
/db_xref="GI:9957182"
/translation="MNLTQSRRLVCLITLIVAAQAQAKVDEAIRPSPDSENL
  SGRYAEGRPSPKLOIDAGVYKGLMYDLNNRGLAEATPVSDVLAISMSR
  VAAVDRSGKRLMOPDYDRGHSRYTCCDAVNRGVALMNGVYTGALDGLILDAK
  TRELMSEQITPAKPYSTGAPRVKKGVIIGNGABRGVDSNADYBELDLYIGNG
  FYTPEDPAQPEYHEPELAATKAPWKGQDQWKLGGGTGVSADYBELDLYIGNG
  SPNREIRSPGGDNLYLSILALPDSKILMHYOTPGETGFTAAQOQLTATLEL
  DGKPRVLMQAPKNGFFVLDLALPDSKILMHYOTPGETGFTAAQOQLTATLEL
  KQVWVWSPSPGAMHVSFNPOTGLMTIPIYOEIDGVYRNGNATPKIDNGCTG
  STHETPRDVAAGALLAMPVORPRAKMRPSPHFWNGGTLTAGVILVQGTADQHA
  YSADKQRLMSFAQGTGYAIPISFSLDEQYVVAAMGCAAPICGDALAPERY
  LSRLLVFKLGGQSLPPLFOATTAIVQRTPAVQASADLDKGLYENISMGCGVG
  AVSGGLIPDLRHTDWRDNFOQIVRDTGLRPLGMPSPNLSIAAEVQIKAYWSRE
  YEDHLKQQAARP"
/gene="pqdA"
/translation="coenzyme PQQ synthesis protein A"
/protein_id="AA09250.1"
/db_xref="GI:9957183"
/translation="WMTKPSYTDLRIGEVTLTYFANR"
/gene="pqdA"
/translation="coenzyme PQQ synthesis protein B"
/protein_id="AA09251.1"
/db_xref="GI:9957184"
/translation="WHIRILGSAAGGFPOMNCNCRGVNAGTLRAQPTQSIAL
  SDGGGHWLGNASPDIRALQLEAFPALQPARQJLDYAGIYVLDQIDHCTGLTIRE
  GCHOVWCTEMHODLTSGPFLNMLSHNGGLQJOLIELDKPTIIPACGLSTAI
  ARSSAPVSPHNGNPGDNLGFLIEDRSRSLFVAPAGVGEHILSWMRADCL
  IYDGTLMRDEMOLCEVQDKLSEMGHLCQSGPGMIVLDBLPSVKILILHINNTF
  ILDLDSPERAELDARGIEVAFDQMSIHL"
```

ORIGIN		Query Match	5.7%	Score 98.8	DB 1	Length 7097
		Best Local Similarity	45.3%	Pred. No. 3.3e-08		
		Matches 702	Conservative	0	Mismatches 807	Indels 39
						Gaps 8
Qy	111	CGCTGTGAATGGAATTAACTACCGCGCGCAACCAAGAAACTATCGCACTCGCCCTGAC				170
Db	3352	CGGAGCGAATGGGTGCTGACGATGCGCGCACCTACGCGGAGACAGCTTTTACGCGCCCTCAA				3411
Qy	171	CCAGATCACTGCGGACCAACCTTGGTCACTTGTGCAACTGTGTGCGCCCGGAGATGGAAGG				230
Db	3412	GCAATATGACGCGGCAACCTGCGCAAGCTGGGCTGGCTGTACTTGGATTTGGAATAA				3471
Qy	231	GG--GGGCGGTACAGGTACAGCGCGATGATCCATGATGCGGTGATGATCTGCAAAACC				287
Db	3472	CAACCGTGGCTTGGAGGCGACCCCGGTGTGTCGACGCGGTCTTACGCGCTGCTGTC				3531
Qy	288	CGGTGATGTGATCCAGGCGCTGGAATGCGCAACAGCGCATGTGATCTGGGAACACCGCC				347
Db	3532	CTGGAGCGGGGTGATGGCCGTGACCTGCGCACGCGCAAGCCGCTTGGCAATTCATCC				3591
Qy	348	CCAACGTCCCGCGGTGCGCCACGCTAAACGCCCAAGCGAC--CGAAGCGCGCGCTGCG				404
Db	3592	CCAGTGAATCGCGGCGCATCCCGCTACCTGTCTGACGCGGTGACCGCGCTGAC				3651
Qy	405	CTTTACGGACGAGCGCTTATTTCAGCTCATGGAACCAATCTGATGCGCTGGAATAT				464
Db	3652	CTGTGGAACGGCAAGGTCTACGTCGCGCGCTGACGCGCGCTGATGCGCTGACGCG				3711
Qy	465	GGAGCGGCGGAGGTGTGATGCAATGCAAGTGAATGCGGCGGAGAACGCGCTTACAG				524
Db	3712	CAGACCGGCG--GCAAGCTGTGAGACGAGACGACCAACCGACCGGCGGCTTACAG				3768
Qy	525	TAAACACCAACGCGCGCGATTTGTGCGCAATGCGGTATGTCGCGGTTTCACTTGCAATA				584
Db	3769	CATCAACGCGGCGCGCGGCGGTGTGAGAGGCAAGTATCATCGGCAAGCGCGCGCGCA				3828
Qy	585	TTGCGCCTTAGATGCTTTATCTCGGCGGACGATTCGCGGACGCGGTGAGAGGTGTGCGG				644
Db	3829	GTAAGCGGTGCGGCGCTTCTTCTCCGCTTACGATGCGGACGCGGCAAGTGTGCTTGGG				3888
Qy	645	CAACCACTTATCCCGGACGCGGCGGAGAGGCTGACGACCTTGGGCGCATGATTTGCA				704
Db	3889	-----TTTCTACAGGTGTGCGGCGGATCCGCGCGGACGCTTACGACACCTGAGTGGC				3942
Qy	705	GCGCGCTGATGACCGCGCTCTGCGGTGACATCACTATGATCCCTGTACGAACTTTGT				764
Db	3943	CGAAGCGCGCAAGCTGGAAGGCGCACGATCTGGAAGCTTGGCGGCGGCGCACCGT				4002
Qy	765	GTTCTATGCTCGACCGCGGTGCGCCAGCGGTCCGAAACCCAGCGCGGACCGCGGCGG				824
Db	4003	GTGGGACAGATGCTTACGACCCGGAAGCTGACCTGTGTACATGCGACCGCGCAACGG				4062
Qy	825	CACGCTGTATGACCAACACCGGCTTTGCGGTGTCCTCCGACACGCGGAGATTGTCTG				884
Db	4063	TTGCGCGTGAACCGGACGATCCGACGTCGCGGCGCGCGGAC--ACCTGACCTGTG				4119
Qy	885	GGCTACCAACCCCTGCGCGGCGGCAACTGCGGCAAGAAATGACGTTGGAATGATGTG				944
Db	4120	GTGATCTTGCGCTTGCGCTGCGGACAGCGGCAAGCTGTCTGTGCACTTACGACACAC				4179
Qy	945	CGCCACGTGATGTGCAACCTTGCGCGGAGATGAGGCTTGGCGGCTCAACCCCA				1004
Db	4180	GGGGAACCTGGGACCTTACCGCCACCCGACGATCACCTTGGCCACCTTGAGGTGGA				4239
Qy	1005	TGCGGCGACCGGCGAGCGCTGTCTGTGACGCGGTGCGCTTGCACACCGGCGACGATG				1064
Db	4240	CGGCAACCGCGCAAGGTGTGATGAGGCGC-----GAAAGACGCTTCTTCTA				4290
Qy	1065	GTGCTTGAATGCGGCGCTCGGCGCAATTCCTGTGCGCGCGGTATACCACTACCAATAT				1124
Db	4291	TGTGCTGATGCGGCGACCGGCGACCTGCTCTGCGGGAAGAAAGTTGCGGACGATGAC				4350

[illegible]

source	FEATURES
gene	<p>Location/Qualifiers</p> <p>1..302101</p> <p>/organism="Pseudomonas putida KT2440"</p> <p>/mol_type="genomic DNA"</p> <p>/strain="KT2440"</p> <p>/db_xref="taxon:160488"</p> <p>200..1066</p> <p>/locus_tag="PP2639"</p> <p>200..1066</p> <p>/locus_tag="PP2639"</p> <p>/note="identified by match to Pfam protein family HMM PF00701"</p>
gene	<p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="dihydrodipicolinate synthase, putative"</p> <p>/protein_id="AN688247.1"</p> <p>/db_xref="GI:24984217"</p> <p>/translaction="WSNFRGIWIALYTPMFRANEIDPPGLEKLYKLLIDGIVAGVCCCTTGAALSKAEQALVDALWAVEGFRNWMGLSGNIRELLAFPAETIQDIDGILLWPCPYIRPQAGIEAFPAITVADAASVPIVYDIPTVYRIERBELRIIVPPIIAWKGCDSEITTMALIEDGQAVLAGEDLOIQFNMLCIGGAGIASAHAYRADIYVMVQCVDSDMAWAGTGFQQLPWIKVAFAPNPNAVKALQIQSLADELREPMQTCNTRFDKLSVYCSIGA"</p> <p>complement(1220..1678)</p> <p>/locus_tag="PP2640"</p> <p>complement(1220..1678)</p> <p>/locus_tag="PP2640"</p> <p>/note="similar to GB:M23326, GB:X15261, GB:X13954, PID:1049135, PID:2358067, PID:312412, PID:37046, PID:37048, PID:37298, PID:37324, and PID:540657; identified by sequence similarity; putative"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="acetyltransferase, GNAT family"</p> <p>/protein_id="AN688248.1"</p> <p>/db_xref="GI:24984218"</p> <p>/translaction="MQYRRTITPDDLPQALCMDAPMOAVAPTLSAGIEFTFOVSAPQARERQGNMLACFVBSGALIGLLEKGRHIMLFLPDELQIGLGRKLMAALEHASEAVTVKASISSVPAYQRYGFTLAGVSGFAGLVYQPEMKRLTPQA"</p> <p>complement(1117..3516)</p> <p>/locus_tag="PP2641"</p> <p>complement(1117..3516)</p> <p>/locus_tag="PP2641"</p> <p>/note="similar to SP:P37911, GB:X07469, SP:P08149, GB:M32091, GB:X07468, GB:X07470, GB:X53633, PID:150279, PID:44911, PID:44913, PID:44915, and PID:509155; identified by sequence similarity; putative"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="iron-sulfur cluster-binding protein"</p> <p>/protein_id="AN688249.1"</p> <p>/db_xref="GI:24984219"</p> <p>/translaction="MKPKYLEVILVDYENNRNRSICPEVLNQSRRVIELSPGKTFEASQELLDAMLASGLVPFPCSRGACSCCYKVSQCHODKQSDADTPPSYPLADEKDEKQSLHSGHSDMRLEIPGMSLDAPLETOAGVKGALSIYELVACQLEFACGAGLQVYMRQGLVSRPCESTIANLPARQGLVTHQVKEGSLPTELLPTLQAGTVLKEGPIGATQWRHDDRPILIPATGTGAGIKKPLLALAGADNLTLYGSSSPAPCPREPLIDVSRVPHFRMPQVLSAQRITQVALSOTHRMERTQYVAGNMLARSQAEQCLAGAVPPIRFAEAPVASGALTTQASSANLHPELEKVPRLSLOGMALARSQRAVAALASQLQVGTGQALEMAHTLOAGMSHTWHPTVIRFGDVTATPQGI:DLQRLVITDILQVDDVPWMDGYEGDITFVGGQALHRAVCYKALHEVFTBQAMGRGIGTVELYVFAESQAKGMOLERNLAGRIADPPHVLVYGDQKLAIEVIVSEVWVLEIQLPTEPVGAFEDDILANASRGDLPQDLTALEPVG"</p> <p>3517..5007</p> <p>/locus_tag="PP2642"</p> <p>3517..5007</p> <p>/locus_tag="PP2642"</p> <p>/note="identified by match to Pfam protein family HMM PF00155"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="transcriptional regulator, GntR family"</p> <p>/protein_id="AN688250.1"</p>

/db_xref="GI:24984220"
/translation="MYQKRWQPLRIDQAQSYQYVEGLVAAISSEGLRPTLP
TRMAQLLVNKRVTIYLAEBATKGLQSVQRGTPVQQLAPGTAAAGVRFAL
ALPEBPAAVYFKANEQAVAOQRPALFPDNGACHRLPAVLRHYRANLRSST
NSVYHSECSRYLSALADMLRHNSLVGAHEHLCIGVQMSLSLVAYLKSGD
VYVERLYPAMEIFROLGARLYVELDHGEGVDALCRQHKRMVYTPHQP
TVYSLAGROOLLELROHAFQVVEDYDEYVQDPLPLASDROORHIVYGL
SKYISPPRCVSYVAVPYVTYLERAMLTIGQPAVMOMLADLINDGELKKHRY
NREYRRRETTLGGCLDPAFGEQISVQAPGGIALVRFPADEINVDQLKALERLVV
RSQRQSPFGHAENLRIGFASLDDETRQATORLDAKAKIGKAR"
5229. .6881
/locus_tag="PP2643"
5229. .6881
/locus_tag="PP2643"
/note="Similar to GP:9947573; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="methyl-accepting chemotaxis transducer"
/protein_id="AA068251.1"
/db_xref="GI:24984221"
/translation="MYPTSTAKMLANKIRQGMFVYLSFSLTILFSTASMMALG
SDQITLEDQTAHQSDRLNNALMIRSRANVSGFIBQLGHDSDAGKRALSTELN
NKSQALVDFEVENAREPALRGATLQATFAVAAVAGQKATQKSLDEQFKVNSD
ACVAMGRLOTLRQOLVTLTISERGQIMESDRRLARAOQLSLGLGVTVVLCWAF
IAQRVHPHREAGGHFRRIASGDLSPVQGGNNEIGQLFHLQMOOSORDLQOIN
NCARQDAATALNAVTEESANMLRQOGELOAAVAVTEMTAVEVARNALITQSC
TSESNLOAOSRROVSENI DGTENMTREIOTSAMLOQGVGROIQVLEIRVSE
QNTLLALNAIRAPAGAGRGAPAVADPRLTAQTOGSEIOMIGSVQAGTEAA
VSMQASTRAQSTLDVTLASGOVLEGTSAIGENERLVVLSAAEEOAQVAREVDR
NLNITRELSNHSAGAQOTSEASKLSGLVGMTALVGRFKV"
7131. .7277
/locus_tag="PP2644"
7131. .7277
/locus_tag="PP2644"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA068252.1"
/db_xref="GI:24984222"
/translation="MCRDLPPVLAGMRQAEHPHIRFTSPACVFCAAAGPRLAAS
ARLP"
7575. .10337
/gene="mgcB"
/locus_tag="PP2645"
7575. .10337
/gene="mgcB"
/locus_tag="PP2645"
/note="Similar to GB:X60459, SP:P17181, PID:1247459,
PID:1247461, PID:1567383, PID:1567385, PID:306914, and
PID:32672; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="magnesium-translocating P-type ATPase"
/protein_id="AA068253.1"
/db_xref="GI:24984223"
/translation="MKRLCVPPQROEHSNMTVKRNTSSAKARETSLSTRAAEAR
NGIAVTLANDISEOGLTEHEAKRLVRGANQVADHPHPLVOLLRALNPFYVL
LTLAGISFVTDVYLPVYAGADADLTXYIIMTVSLSLRFQERSNKAALAK
AMRTATVLRQICQAPRLREVPMDLVAQIVQLSAGDMI PADILULEARDLIS
QAVLTGALPVEKYDTLGNVAKSAEHAQODNLELPNICFGNTVINGGAARVVA
AGRTRTFGSLAKAIGRSOTAQFDRGVNSVSLIRPLVWVVPVPMINGVAKDMA
DAFLALAVAGLTPEMLPMIVSANAKAVAMARKVYVVKLAINQNGMDVYCTD
KCTGTQDRIILFHHYGPQGDYKRIILSLAMNSHQSGIRLMOAYLHPGQDHCP
QAVYAYKVDLEPFDTRRRLSVVNSALGDHLVSKGAEVEMLAIAHVQGDVVA
LDCRQOOLMARVADANQDFRVLVATVQIPADBGKQVHATEDRDLVQLGLFTLD
LPKETAQPAIALRDMGVQVFLVATVTSKVRCEGLAGQPLQODLEGMDT
LKQVEERTVFAKLTPLQKSRVYKALQANGHTYFGGGINAALRADVAGISDSG
TDIAKESADIIILERSIMYLBEGVLKRGTFGNIMKYLCTSSNFGVYVLSVAF
IPRLPMLATHILQNTMVDQSLSLPMDSMDKEPLSKPKMDARVIGGPMIGTSS
IDPITTPALMWTYRANVEMQALFQSGHPIGSLISQTLVYMLATRYKVPFQOSRAL
PVVLATGLVVALGITYIPSPVGAMGLVFLPWEYFPMVATLLGICVVAQAQAKTLYIR

gene
10341. .10735
/locus_tag="PP2646"
10341. .10735
/locus_tag="PP2646"
/note="Identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AA068254.1"
/db_xref="GI:24984224"
/translation="MQGEQTMRLICAGRNVAUTRRCQALDDVQRFQPIRVLIHGS
QFLGGEISWREHGHADLVYPPNMOJHGQERLRNLMINDSRDLIALFGGDT
EELIARARCAELGVVYAROPCEPDDCS"
complement (10843. .12069)
/locus_tag="PP2647"
complement (10843. .12069)
/locus_tag="PP2647"
/note="Similar to GP:5738822; identified by sequence
similarity; putative"
/codon_start=1
CDS
gene
complement (10843. .12069)
/locus_tag="PP2647"
complement (10843. .12069)
/locus_tag="PP2647"
/note="Similar to GP:5738822; identified by sequence
similarity; putative"
/codon_start=1
Query Match 5.7%; Score 98.4; DB 1; Length 302101;
Best Local Similarity 45.3%; Pred. No. 2.6e-08;
Matches 718; Conservative 0; Mismatches 821; Indels 47; Gaps 8;
125 TTAACACGCGCGCAACCAAGAAACCTATCGCACTCGCCCTGACCCAGATCACTGCC 184
DB TGACCAATGCGCTGGCGGTGACAGGGCCAGCGCTACAGCCCGCTGACCTGCTCAATGTCA 44035
QY 185 ACAACGTTGGCTCACTTGCACCTGCTCGGC-----CGCGGGAGTGAAGCGGGGG 225
DB 44036 ATACGTCAGAGACCTCGCGCGGTCTGGCGCTTCTTCCCTGGCGGAGAACACAGCG 44095
QY 236 CCGTACAGTCAACCGCATGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 295
DB 44096 CCGTACAGTCAACCGCATGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 44155
QY 296 TGATCAGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 355
DB 44156 GGGTGTTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44215
QY 356 CCGCGTTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 415
DB 44216 ---CGATGATTCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 44272
QY 416 CGACCTCTATTTCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 475
DB 44273 ACTGATGATTCCTTGGCGACCGCTGACCGCAAGCTGATGATGATGATGATGATGAT 44332
QY 476 AGTCTGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 535
DB 44333 AGTCTGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44389
QY 536 GACCGATTCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 595
DB 44390 CGCGATGATTCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44449
QY 596 GATCTTATTCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 655
DB 44450 TGGGCAAGATTCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44509
QY 656 TCCCGCAGCGCGGCAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 715
DB 44510 TCGAAGGCAACATGCGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 44568
QY 716 TGAACGCGCTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 775
DB 44569 GCGCGTGAAGCGGCGCAAGACCTGCGCTGCGCACTGTGGAAGACCGCGCGCGCG 44628
QY 776 CGACCGCGCTGCGGCGCGCAAGACCTGCGCTGCGCACTGTGGAAGACCGCGCGCGCG 827
DB 44629 TGGCTGGGCGCTTACTGACACCTGGAACCAACCTGATCTGTTGTTGTTGTTGTTGTT 44688


```

QY      828 GCTGATGACACAAACCCGCTTTGGCGTGCCTCCGACACGGGCGAGATTGCTGCGC 887
Db      44689 GCGCGGTGAACCTGCACTTCCGCGCGGTGACAACTGTAATCTCTCTCACGCTTGCA 44748
QY      888 T--CACCAGACCTCGCGCGGACAACTGGGACCAAGAATGCAAGTTGCA-----G 936
Db      44749 CTGAACCCGAGACGAGGACCATCAATGAGCATTTCCAGACGCGCATGACGGCTGG 44808
QY      937 ATGATGTCGCAACGTCGATGTGCAACCCCTCGGCGAGATGAGAGGCTTGGCGGCATC 996
Db      44809 GACTTCGACGCGCTCAAGAGCTATCTCTTCAACTACAGAGCGCGGCGAAGAGCTC 44868
QY      997 AACCCCAATGCGGCGACGGGCGAGCGCGCTGTGCTGAAGGCTGCGCTTGCAGAACGGC 1056
Db      44869 AAGGCTCCGCGACGAGACCGGACAGGTTCTTCACTGCTGCAAGCCACCAACGCGC 44928
QY      1057 ACATGATGATGCTTTGATGCGGCGCTCGGGGGAATTCCTGTTGGGCGGTATACCAACTAC 1116
Db      44929 AAGTTCAATCCGCGGCTTCCCTTCTGTGACAAAGTCACTTGGGCTCATGCTGAGACAG 44988
QY      1117 ACCAATATGATCGCTCGATGACGAGACGGCGCTTGTGAAGGTGAACGAGATGCGGTG 1176
Db      44989 GACGCGCGCGCGATCTCAACAGACGCGCGCGCGCGACACCGGCGACGAGGCCAAG 45048
QY      1177 CTGAAGAAGCTGGAACGTTGATGATGAGCTCTGCGCGCATTTCTCTGGGTGGGCGGACTGG 1236
Db      45049 GGCAGCTCGGTG-----TTGCTGCGCGCGCGCTTCTCTGCGCGCAAGAACTGG 45096
QY      1237 TCGTCAGCGCGCATGAAACCCGACACCGGCGCATTTACTTCTTGGCGCGTGAACAAATGCTGC 1296
Db      45097 ATGCGCATGCGCTCAACAAAGAGACAGAGGCTGTTCTACGTCGCGCTCAACAGAGTGGGCGC 45156
QY      1297 TACCATTTATGCGCGCTTGTATCAAGATTGAGCGCGCTGACGCTTATACACCAAGCGCG 1356
Db      45157 ATGACATCTGGAACGAAAGCGATCGCTATAGAAGGTGCGCGGTCTCTGCGTGGCGCGC 45216
QY      1357 ACCGCAAACTCGCGCGCGCTTTGAATAATGAGCGCGCATTCAGCGGATGATGATCAGC 1416
Db      45217 TTTCACATCAAGCGCTCATGAAAGTCAATCATCGGCTGCTGCGCGCATGACCGGCTC 45276
QY      1417 ACCGCGCGCACCTTGTGCGCGCGGCGCGCTGCGCGCAACTACCTGCGCGCTTTGTG 1476
Db      45277 AGCGCGAAGGAGTGTGCGCGCGCAACAAGACTATCGCGCTGCGCGCGGTGTGCTGAC 45336
QY      1477 ACCGCAAGCGGTGTGCTTCAACGCGCGGACGAGCGGATTTCCGTGCGCTCAAGCAG 1536
Db      45337 ACCAAGGCAACTGTTGTGTCAGCGCGGACGCGGATTTCCGTGCGGATTTCAACGCGC 45396
QY      1537 GAAACCGCGGACCTTTGTGTCAGCGCGGCTTGTGCGAGCGGTGCGGACGCGGCGATC 1596
Db      45397 AAGACCGCGGACAGTGTGTCAGCGCGGCTTGTGCGAGCGGCTTGTGCGGCTGCGCGCTC 45456
QY      1597 AGCTACAGATTGACGCGCGTGCATATATGCGCATTCGTCGCGCGGTCTGACCTATGCTC 1656
Db      45457 ACCGCGAAGGAGCGCGCGGACATACGTTTCGCTGCTGCGCGCGCGCGCGGTG 45516
QY      1657 ACCGCAATTGACGCGCGCTGCGCGA 1682
Db      45517 CCGCTGTGCGCGCGCGGAGAGTGGCAA 45542

```

RESULT 14
AY316747.0/c
WPCOMMENT

Sequence split into 4 fragments LOCUS AY316747 Accession AY316747

Fragment Name Begin End

AY316747_0 1 110000
AY316747_1 100001 210000
AY316747_2 200001 310000
AY316747_3 300001 357655

LOCUS AY316747 357655 bp DNA linear BCT 02-JAN-2004
DEFINITION Rhizobium sp. NGR234 megaplasmid 2 contig 1, complete sequence.

```

ACCESSION AY316747 GI:36958823
VERSION AY316747.1
KEYWORDS
SOURCE
ORGANISM Rhizobium sp. NGR234
REFERENCE Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
AUTHORS 1 (bases 1 to 357655)
Broughton, W.J., Perret, X., Staehelin, C., Schmitz, R.A., Raasch, C., Liesegang, H., Gottschalk, G. and Stett, W.R.
TITLE Comparative DNA analysis of two large contigs of the Rhizobium sp. NGR234 megaplasmid 2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 357655)
Broughton, W.J., Perret, X., Staehelin, C., Schmitz, R.A., Raasch, C., Liesegang, H., Gottschalk, G. and Stett, W.R.
AUTHORS Direct Submission
TITLE Submitted (06-JUN-2003) Institut fuer Mikrobiologie und Genetik und Laboratorium fuer Genomanalyse der Universitaet Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
FEATURES
source
1..357655
/organism="Rhizobium sp. NGR234"
/mol_type="genomic DNA"
/strain="NGR234"
/db_xref="taxon:394"
/clone="contig 1"
/plasmid="megaplasmid 2"
complement(1..1449)
/locus_tag="RNGR00476"
complement(1..1449)
/locus_tag="RNGR00476"
/codon_start=1
/transl_table=1
/product="Putative sensor histidine kinase protein"
/protein_id="AAQ87601.1"
/db_xref="GI:36959176"
/translation="MTRISDMILDPADNPENSTAKLTIAEVLNROVQASGEKVAYEFOFPAALIEDVRQRTSLSTKLINETNAPLAENRATERRARNLADILEAEGFALPDASILWYNRPFCDLEDRATIRPGLSPADYIEACSRRLAPGGGADNNAIRNQRRTSVFQVALAGRWLOVSQPTSDGTVLQTDVEIIMERSGERMLDDQARLIATLEHINQICIPDAARLVGWRERLRLDILVTYLRTEGTGDTLQRMORDGELHSITPELPSFSAVOCTQRPPLFELIRTSGYMLVFVFAETPDGTFVMSFSDVTRERLTAQAMQANASLEAVVAATLEDRLAALDAERANSRTRAFVAAASHDLOPLSAKLFVNASQDETERAGATLDRKALASVERALIGALLDISLDAVAVEISVRLAPLQRLTDERPAPMERKLOLATLPCALTYFS"
complement(1446..2627)
/locus_tag="RNGR00477"
complement(1446..2627)
/locus_tag="RNGR00477"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAQ87249.1"
/db_xref="GI:36958824"
/translation="MTRIGSGGVVDAIVPSASTANDPNAETLSRLQAGPRLIMLFISPAADLERLPAAARVFDVPIVCTTAGEISGAGYDGTVALALPSEFRVDTVTIPMLRDPDPAALAAVRAVRVAVRYPALAHFALLVDGSKDELTLSLAAGLGAMPILGSGAGDGTAVRTFVLDGGRSRNAALVALVRECPVRFVDFHMTERRMVTNADPRRVRVQINAPRAHVEYARLVGDPQOLSPTFRAHVRVAKMTGHHVRAIRQVNDNDLVFSAIDEGVYLTTPEDMNGQLRELAGCMERFPAILAEPCLIRRLAEKQVIGRVSGILRRHGVGVSTYGEQFPHVAVQMTGVALIPGTVLPKQVPE"
complement(2701..3423)
/locus_tag="RNGR00478"
complement(2701..3423)
/locus_tag="RNGR00478"
/codon_start=1
/transl_table=1
/product="Transcriptional regulatory protein"
/protein_id="AAQ87250.1"
/db_xref="GI:36958825"
/translation="MIIIRNSFQMPRTISVPQPLFKPVELRSALIVDDHPLFCALAM

```


QY 509 AAGACGGCTTACACAGTAAACCAACGAGGCGCATTTGTCGCAATGCGCATTCGTGCGG 568
 DB 4778 AGGAAGGCT---ACAGTACACGCGCGCGCTTGATCGTGAACGGCGCTCATATCAACGG 4722
 QY 569 GTTCCACCTGACCAATATTCGCCCTATGAGATGCTTTATCTGCGGAGGCAAGATTCGGCAGCG 628
 DB 4721 GCAATTCGGGTGCGAGTTGCGTATCGTCGCGAGGATTCAGAGCCGCGCATGCGCAAGACCG 4662
 QY 629 GTGAGAGCTGTGAGCGCAACCACTTTATCCGAGCGCGGCGAAGAGGATGACGACATT 688
 DB 4661 GCGAATCGTATGAGACCCGCTCGGTCATGAGAGGCAAGTCGGGATTTCAAGGGCAAG 4602
 QY 689 GGGGCAATGATTTTCGAGGCGCGCTGATGACCG---CGTCTGAGGATCAATCACTATG 745
 DB 4601 ATAGCTCGATGACCGGAGCCGTGAATGCGACCTGGCCCGCGCATCTTTGGAAGACAGCG 4542
 QY 746 ATCCCGTGAAGCACTTGTGTTCTATGAGCTCGACCGCGCGTGGCCCGAGCGTCCGAAGCC 805
 DB 4541 GCGGCGCCACCTGCGCTCGCGGATCTTACAGCGCCATACCGATACGCTGCTTTCGGCA 4482
 QY 806 AGCGCGACAGCGCGGCGCGACGCTGTATGACCAACACCGCTTTGCGGTGCGTCCG 865
 DB 4481 CCGGCAACCCGCTCTTTGGAACAGCCATCTGCGCAACCGCGCAAGCGGTGAGGGCA 4422
 QY 866 ACAAGCGCGAGATTTGCT-----GGCGTACCAAGACCTTCGCGCGCGCAACTGAGGAC 919
 DB 4421 ACAAGGCGCAACCTCTACCGCGCGAGCGCTCGGCATCAATCCGAGGAATGCGGAGA 4362
 QY 920 AAGATGCAAGCTTCGAGATGATGTCGCCA-----CGTCAATGTCACACCT 967
 DB 4361 TCAAGTGCACCTTCCAGACGACCCGAGGAGGAGCTGGGATTTGACGGGTGGAATGAG 4302
 QY 968 CGGCGGAGATGAGAGGCTTCGCGCCATCAACCCCATGCGGAGAGCGCGCGAGCCGTG 1027
 DB 4301 TCGTGCCTTCATCTCCAAAGAGGAGGACGAAAGCGCTTTCGCAACCGCGCAAGCGCT 4242
 QY 1028 TGTGACGAGGTGCGCTTTCGCAAGCGGACGATGTCGTGTTGATGCGGCTCGGCG 1087
 DB 4241 TCTTTCACGTTCTCAACCGGAGAGGAAAGTTCTCCGCGCTATCTTTTGTTAAGA 4182
 QY 1088 AATTCTGTGAGCGCGGTATCAACACTCAACCAATATGATGCGTCGATGACGAGACCG 1147
 DB 4181 ACATCACTGCGGAGAGGATGACGAGACCGGTGCGCGCATCTATGCGGAGAAACC 4122
 QY 1148 GCCTTGACCGTGAACGAGATGCGGTGCTGAAAGCGTGAATGACGCT 1207
 DB 4121 GCGCCGCGCATTCGTTCTTCGCGCGCGACGCGCGCAAGGGGCGAGGCTTTTCGCT-- 4064
 QY 1208 GCGCGACCTTCTGAGGTGCGGCGGACTGTGTCGACCGCGCATGAAACCGCGACCGCGCA 1267
 DB 4063 -TCCCTCTCTCTCGCGCGGCAAGACTGAGTGCAGTGCCTACAGCGAAGAGGAGAC 4005
 QY 1268 TTATCTTCTTTCGCGCGTGAACAAATGCTGCTCAATATATGAGCGCTTGAATCAAGATT 1327
 DB 4004 TCTTTTACGCGCTCGAAGCAATGAGGATGATATCTGGAATGAGCCGATCTCTTCA 3945
 QY 1328 GCGCGCTCGACGTCTATTAACACAGCGCGCACCGCAAACTCGCGCGCGGCTTTGAATA 1387
 DB 3944 AGAAGGGGGCGGCTTCTCTCGGGCGGCTTCACTCAAGCGCTGTTTCAAGACTTA 3885
 QY 1388 TGGGCGCGATCGACGCGATTTGATATCAAGACCGGGCGCATCTTGTGTGCGGAGCGCC 1447
 DB 3884 TCGGCTCGCGAAGGCGATCGACCGCAACAGCGCGAGATCAAGTGGGAATCAAGAAAG 3825
 QY 1448 CTGCGGGAACCTACTCGCGCGCTTTTTCGACGCGGACGCGGTGTGTTCAACGCGGGA 1507
 DB 3824 CCGCGCGCGCTGTGGGCGGTGTATGCGCATCTGCGGGGGCGTGTCTTCGGAAGCG 3765
 QY 1508 CCGACCGCTATTTTCGCTGCGCTCAAGCGGAGAAACCGGCGAGACTTTGTGCGAGCGCGCTC 1567
 DB 3764 CCGAGGCGGATTCAGAGCGCTCGACGAGAGACCGGCGAAGAGCTGTGCTTTCAG 3705

QY 1568 TTGCGACGCTCGCGACGCGGCGAGCGATCAAGTTCGACGCGCGCTGCAATATATCG 1627
 DB 3704 CCGGCTCGGCGGTGATCGCGCCAGCGATCACTGAGAGACGAGCGGAGCAATATGTGT 3645
 QY 1628 CCATGCGTGGCGGCGGTGACCTATGACGACGCAATTGAACCGCCGCTGCGCGAGCA 1687
 DB 3644 CGGTATCTCGGTTGGGGCGCGGCTCCCTTTGGGGGCGGAGATCCCAAGAG 3585
 QY 1688 TCGATTGACCTCGGTGCGTATGCGATCTATGCTTTGACCTGCGCGAGTA 1739
 DB 3584 TCAATATCTCAACCAAGTGGCATGTTGTGACCTTCGCTTCGGAAGA 3533

RESULT 15
 AF355798 2550 bp DNA linear BCT 14-MAR-2002
 LOCUS Pseudomonas butanovora 1-butanol dehydrogenase BDH precursor, gene,
 DEFINITION complete cds.
 AF355798.2 GI:16579769

ACCESSION
 VERSION AF355798.2
 KEYWORDS
 SOURCE Pseudomonas butanovora
 ORGANISM Pseudomonas butanovora
 Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales;
 Rhodocyclaceae; Thaueria.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 1 (bases 1 to 2550)
 Vangnai,A.S., Arp,D.J. and Sayavedra-Soto,L.A.
 Two distinct alcohol dehydrogenases participate in butane
 metabolism by Pseudomonas butanovora
 J. Bacteriol. 184 (7), 1916-1924 (2002)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 2 (bases 1 to 2550)
 Vangnai,A.S., Arp,D.J. and Sayavedra-Soto,L.A.
 Direct Submission
 Submitted (02-MAR-2001) Botany and Plant Pathology, Oregon State
 University, Corvallis Hall, Corvallis, OR 97331, USA
 3 (bases 1 to 2550)
 Vangnai,A.S., Arp,D.J. and Sayavedra-Soto,L.A.
 Direct Submission
 Submitted (01-NOV-2001) Botany and Plant Pathology, Oregon State
 University, Corvallis Hall, Corvallis, OR 97331, USA

REMARK
 COMMENT
 FEATURES
 On Nov 1, 2001 this sequence version replaced gi:13448822.
 Location/Qualifiers

source
 1. 2550
 /organism="Pseudomonas butanovora"
 /mol_type="genomic DNA"
 /db_xref="taxon:86174"
 369..2444
 /codon_start=1
 /transl_table=11
 /product="1-butanol dehydrogenase BDH precursor"
 /protein_id="AAK27220.2"
 /db_xref="GI:16579770"

CDS
 16579770..16579770
 /translation="MLTTFPARKREESVPLRKIGORALLGLSCVLTSTTSFAAGGEMR
 THGYDDAGTYSPLAQITPDNAKELIWMVSVDLSSRGVATPIVVDGMYVIAFMVY
 VHALDVSGKRLMTYDEPVRKKNACVVNAGVAVHAGVFGSLDGLVALIDAR
 TGRVWRNTLLIDDKRYTITGAPRVYKGVVINGGAEKGVGYIATPPTAASRG
 VVPRGDDPSLPEPDAWEAAKTMDPAQVGLSGRGPGRVGLDGLVPRKAGPCTTSAR
 TPSPMSRRKSPRAGDNLTTASIVALEPDGTBYMHQCPADRWDTSTODLILADI
 ELGGKPRKVLHAKNGFFVIDRTDKTISAQIFVVMNATGIDENGRLENBEGAM
 PGLHSMPFAPASRTNHSYSPTQSAFEPQNTPLVIOEDKMSYNOAQOQAA
 GIGNLLMLVAPRPAQPFGRLLAMPVOQKEVWRKEHVS PMNGTLLVTAGNVVFG
 TADRLALFAPADKELMSAPMGTVIAAVTYEVDKQVSLAVGGVYGNVTRAS
 ERRPTGVYTFALGKRAMPAFTFVOLNNVSGVDVYPDVAVSGTGLYTNVCFCHG
 PCNDKGGGIPNLGISTATIAHLDQFVFKRPFMRGNPDPDTKLTLPQVEXIKAFILG
 TADVVRPK"

ORIGIN
 Query Match 5.4%; Score 94.8; DB 1; Length 2550;
 Best Local Similarity 48.3%; Pred. No. 2e-07;
 Matches 452; Conservative 0; Mismatches 442; Indels 42; Gaps 5;

QY	105	CCCGCCCGCTGATGAATGATTAATTAATCAAGCCGCGCAACCAAAAACTATCCGCACTGCC	164
Db	479	CGCTGCCGCTGCGCAATGCGCGCACGCAATGCTTATGACACCCCGCGCAACGCGCTTACAGCC	538
QY	165	CCTGACCCAGATCACTGCGCGCAACAGCTTGGTCTGATGTCACATCGCTCTGCGCGCGGAT	224
Db	539	CCTGCGCCAGATCAACGCGCGCAACATGCCAAGAACTCGGCGCTGGCTTGTGCTTCAAGCCT	598
QY	225	GGAAGCGCGGG---GCCGTACAGGTACAGCGCGCATGATCATGATGAGCGCGATGTATCTGAC	281
Db	599	GGAATCTGTCGCGCGCGCGCTGCAAGCCACGCGCCATCTGCTGATGCGCGGATGTACGTAC	658
QY	282	AAACCCCGGTGATGTGATCAAGCGCGCTGGAATGCGCAAAACAGCGCATCTGATCTGGAACA	341
Db	659	CGCGCGCTGGAAGCGCTGTGATGCGCTGGAAGTGGCGCAGCGGCAAGCGGTGTGTGACCTA	718
QY	342	CCGCGCGCCACATGCGCGCGCGCTGCCACGCTTAAAGCGCCAAAGCGAC---GCCAAGCGCG	398
Db	719	CGACCCGGAAGTTCGCGCGCGAAGAGGCGCAAGATCTCTTGTTCGACGCTGTGAAACCGCG	778
QY	399	CGTCGCGCTTTACGCGACAGAGCTCTTATTTCACTCATGGAACAACCATCTGATCGCGCT	458
Db	779	CGTGGCGGTGATGAAGGCGCAAGGTCTTGTGTCGCTCGCTCGACGCGCGGCTGATGCAT	838
QY	459	GGATATGAGAGACGGGCGCGAGTCTGATTTGGATGTGCAACGTGATGCGCGCGAAGACGCTT	518
Db	839	CGACCGCGAGACCGCGCAAGCGGCTGTGGAAGCGCAACGCTGATCGACGACGACAGCGCC	898
QY	519	GACCAGTAAACCAACAGGGGCGCATTTGTGCGCAATGGCGCTCATCTGTGCGGGTTCCACTG	578
Db	899	CTACACCATCAACCGGTGCGCGCGCGGTATCAAGGGCAAGGTCGTCATGCGCAACGCGCG	958
QY	579	CCAAATATTCGCCCTATGGAATGCTTTATCTCGGGGACGA-----	617
Db	959	TGCGGAGTTCGCGCGTGCAGCGGTACATCAACGCTTACGACCGGACGCGCGAGCAGGCC	1018
QY	618	-----TTCCGCGACGAGTGAAGAGCTGTGCGCGCAACCATTTATCC-----GCAGC	664
Db	1019	TGGCGGTGTTCCCGGCGCGGGTGATCCGTCGTGCGCTTGAAGATGCTCTGATGGAAGGC	1078
QY	665	CGGCGGAAGAGGTTACGAGACTTGGGCAATGTTTGAAGCGCGCTGGAATGACCGGCG	724
Db	1079	AGCGCGCAAGACTCTGGAACCGCGCGCGCAAGTATGGAAGACGCGCGCGCGGACGCG	1138
QY	725	TCTGGGTGACATCACTATGATCCCGGACGAACCTTGTTCTAT--GGCTGCACGG	782
Db	1139	TGTGGAATCTGATGTTTATACCGGAAGCTGATTTCTGTGTACATGGCGACCGCGAC	1198
QY	783	CGTGGGCCAAGCTCCGAAACCCAGACGCGGCAACGCGGCGCGGACCGTGTATGCGACCA	842
Db	1199	CCCTTCGCGCTGTGCGCACCGCGAAGCTGATGACCGCGCGCGCGGACCAACTGTACACCGC	1258
QY	843	CACCCGCTTTGGCGGTGCTCCGACAGCGGGAGATTTGCTGGGCTTCCACGACCTTGC	902
Db	1259	GTCACTCGTCGCGCTGCGCGCGGACACCGGCGAATAGTCTGACCTTACAGCAGACGCC	1318
QY	903	GCGCGACAACCTGGGACCAAGATGCACTTTCGAGATGATGTGTGCGCAACGTGATGTGCA	962
Db	1319	GCGCGACAACCTGGGACTTACACTCTGACGAGGACTGATCTCGGCGGACATGCAACTGG	1378
QY	963	ACCTTCGCGCGAGATGAGAGGCTTTCGCGCATTAA	998
Db	1379	GCGGCAAGCGCGGCAAGTATCTTCATGCGCCCA	1414

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 13:16:29 ; Search time 942 Seconds
(without alignments)
9696.383 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1740
Sequence: 1 atgaaccacacacagctgct.....tcttcgactgcgcagctaa 1740

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: N_Geneseq_23Sep04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2002bs:*
9: geneseqn2003as:*
10: geneseqn2003bs:*
11: geneseqn2003cs:*
12: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1740	100.0	1740	2	AAV29054	AAV29054 Alcohol a
2	1740	100.0	1740	12	ADN10955	ADN10955 Gluconoba
3	1206.2	69.3	1743	12	AD134117	AD134117 Ketogulon
4	1206.2	69.3	1743	12	AD134117	AD134117 Ketogulon
5	1130.2	65.0	1740	2	AAV29051	AAV29051 Alcohol a
6	1079.8	62.1	1737	2	AAV29053	AAV29053 Alcohol a
7	1071.2	61.6	1737	12	AD134114	AD134114 Ketogulon
8	1071.2	61.6	1737	12	AD134114	AD134114 Ketogulon
9	1053.6	60.6	1740	2	AAV29052	AAV29052 Alcohol a
10	1052	60.5	1740	12	AD134116	AD134116 Ketogulon
11	1052	60.5	1740	12	AD134116	AD134116 Ketogulon
12	625.8	36.0	1827	10	AD594132	AD594132 Alcohol/a
13	88.8	5.1	2214	2	AAQ13580	AAQ13580 A.alcoacet
14	87.2	5.0	2214	2	AAQ20383	AAQ20383 ADH comp1
15	85.6	4.9	2265	3	ABA96892	ABA96892 Gluconoba
16	85.6	4.9	2265	3	AAQ31153	AAQ31153 Sorbitol
17	85.6	4.9	4830	3	ABA96894	ABA96894 Gluconoba
18	85.6	4.9	4830	3	AAQ31156	AAQ31156 DNA encod
19	77	4.4	1969	11	ABD16767	ABD16767 Pseudomon
20	77	4.4	2058	11	ABD16858	ABD16858 Pseudomon
21	77	4.4	2637	11	ABD16799	ABD16799 Pseudomon

22	72.6	4.2	510	11	ABD16908	ABD16908 Pseudomon
23	70.4	4.0	2625	4	AAI26354	AAI26354 P. putida
24	64.6	3.7	1050	6	ABQ90330	ABQ90330 M. capsul
25	64.6	3.7	1458	6	ABQ90364	ABQ90364 M. capsul
26	55	3.2	2388	2	AAQ01811	AAQ01811 Nucleotid
27	53.8	3.1	1809	11	ACH99470	ACH99470 Klebsiell
28	53.6	3.1	4412	10	ADD24915	ADD24915 DNA encod
29	52	3.0	5580	8	RCC58392	RCC58392 Gluconic
30	51.4	3.0	2000	8	ADA71938	ADA71938 Rice gene
31	50.6	2.9	2000	8	ADA71938	ADA71938 Rice gene
32	47	2.7	7155	10	AD123900	AD123900 Streptomy
33	47	2.7	37360	10	AD123892	AD123892 Streptomy
34	46.6	2.7	536	10	ADB68842	ADB68842 Minorit
35	46.4	2.7	2244	2	AAT10429	AAT10429 Mature is
36	46.4	2.7	2244	2	AAV23640	AAV23640 Flavobact
37	46.4	2.7	2334	2	AAT10428	AAT10428 Isoamylas
38	46.4	2.7	2334	2	AAV23639	AAV23639 Full leng
39	46	2.6	10982	4	AA546264	AA546264 DNA encod
40	45.8	2.6	109519	5	AA508693	AA508693 Micromono
41	45.6	2.6	358	6	ABN24131	ABN24131 Human ORF
42	45.4	2.6	2000	4	AAI90276	AAI90276 A. nuxar ge
43	44.2	2.5	645	11	ACH99344	ACH99344 Klebsiell
44	43.6	2.5	666	11	ABD17509	ABD17509 Pseudomon
45	43.6	2.5	894	8	ACA37564	ACA37564 Prokaryot

ALIGNMENTS

RESULT 1	AAV29054	AAV29054 standard; cDNA; 1740 BP.
ID	AAV29054	standard; cDNA; 1740 BP.
XX	AAV29054;	
AC	AAV29054;	
XX	10-AUG-1998	(first entry)
DT	10-AUG-1998	(first entry)
XX	Alcohol and/or aldehyde dehydrogenase B	nucleotide sequence.
DE	Alcohol and/or aldehyde dehydrogenase B	enzyme; recombinant organism; aldehyde;
XX	ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulonic acid;	
KM	L-ascorbic; inhibition; ds.	
XX	Gluconobacter oxydans.	
OS	Gluconobacter oxydans.	
XX	Key	Location/Qualifiers
FX	1. 1737	
FT	/*tag= a	
FT	/product= "alcohol and/or aldehyde dehydrogenase B"	
PN	EP832974-A2.	
XX	01-APR-1998.	
PD	11-SEP-1997;	97EP-00115801.
XX	11-SEP-1997;	97EP-00115801.
PR	19-SEP-1996;	96EP-00115001.
XX	(HOPE)	HOFFMANN LA ROGHE & CO AG F.
PA	Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;	
XX	WPI; 1998-1995228/18.	
DR	P-PSDB; AAW37876.	
XX	Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase	
PT	enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-	
PT	sorbitol to 2-keto-L-gulonic acid.	
XX	Disclosure; Page 33-34; 59pp; English.	
PS	This is the nucleotide sequence encoding the Gluconobacter oxydans	
XX	alcohol and/or aldehyde dehydrogenase B enzyme. The enzymes or	

recombinant organisms can be used to convert suitable substrates to aldehydes, ketones or carboxylic acids, especially to convert L-sorbose or D-sorbitol to 2-keto-L-gulononic acid, which can be converted to L-ascorbic acid by standard procedures. The derivatives of AADH enzymes have desired substrate specificity, higher affinity to a substrate, lower affinity to an inhibitory compound, higher stability against temperature and/or pH and higher catalytic speed

Sequence 1740 BP; 340 A; 529 C; 545 G; 326 T; 0 U; 0 Other;

Query Match 100.0%; Score 1740; DB 2; Length 1740;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATGACCCCAACAGCGCTGCTTGCACAGCGCGCCGCTATTTGCTTACCGCGCCGCC 60
Db 1 ATGACCCCAACAGCGCTGCTTGCACAGCGCGCCGCTATTTGCTTACCGCGCCGCC 60
Qy 61 GCATTGGCGGAGTAAACCCCGATTACCGATGAACTGCTGCGAACCCCGCCCTGTGA 120
Db 61 GCATTGGCGGAGTAAACCCCGATTACCGATGAACTGCTGCGAACCCCGCCCTGTGA 120
Qy 121 TGAATTACTACGCGCGCAACCAAGAAACTATCGCACTCGCCCTGACCGAGATCACT 180
Db 121 TGAATTACTACGCGCGCAACCAAGAAACTATCGCACTCGCCCTGACCGAGATCACT 180
Qy 181 GCCGACAACGTGTGCTGCTGCACTGTGCTGCGCGCGGAGTGAAGCGCGCGCGTA 240
Db 181 GCCGACAACGTGTGCTGCTGCACTGTGCTGCGCGCGGAGTGAAGCGCGCGCGTA 240
Qy 241 CAGGTACACCGCGATGATCCATGATGAGCGATGATCTGCAAAACCCCGATGATGATC 300
Db 241 CAGGTACACCGCGATGATCCATGATGAGCGATGATCTGCAAAACCCCGATGATGATC 300
Qy 301 CAGCGCGTGTGATGCGCAACAGCGCATCTGATCTGCGAAACCCCGCGCAACTGCGCGC 360
Db 301 CAGCGCGTGTGATGCGCAACAGCGCATCTGATCTGCGAAACCCCGCGCAACTGCGCGC 360
Qy 361 GTGGCAACGCTTAACCCCAAGCGGACCGCAAGCGGCGTGCCTTTTACCGCAAGAC 420
Db 361 GTGGCAACGCTTAACCCCAAGCGGACCGCAAGCGGCGTGCCTTTTACCGCAAGAC 420
Qy 421 CTCTATTTTCACTCATGAGCAACCACTGATCGCGCTGATATGAGAAGCGGCAAGTC 480
Db 421 CTCTATTTTCACTCATGAGCAACCACTGATCGCGCTGATATGAGAAGCGGCAAGTC 480
Qy 481 GTATTGATGTGCAACGTGATGCGGCAAGACGCTTGAACCAATTAACCAAGCGGCGG 540
Db 481 GTATTGATGTGCAACGTGATGCGGCAAGACGCTTGAACCAATTAACCAAGCGGCGG 540
Qy 541 ATTGTGCGCAATGCGCTGATCGCGGCTTCACTGCAATATTTGCGCTATGATGATC 600
Db 541 ATTGTGCGCAATGCGCTGATCGCGGCTTCACTGCAATATTTGCGCTATGATGATC 600
Qy 601 TTTATCTGCGGCAACGATTCGCGAGCGGTGAGAGCTGTGGCGCAACCACTTTATCCG 660
Db 601 TTTATCTGCGGCAACGATTCGCGAGCGGTGAGAGCTGTGGCGCAACCACTTTATCCG 660
Qy 661 CAGCGCGGCAAGAGGCTGACGACTTGGGCAATGATTTGAGGCGCGCTGATGATC 720
Db 661 CAGCGCGGCAAGAGGCTGACGACTTGGGCAATGATTTGAGGCGCGCTGATGATC 720
Qy 721 GCGGTCTGGGCTGAGATCACTATGATCCGTGACGAACCTTGTCTATGCTGAC 780
Db 721 GCGGTCTGGGCTGAGATCACTATGATCCGTGACGAACCTTGTCTATGCTGAC 780
Qy 781 GCGGTGGGCGCAAGCGTCCGAACCCAGCGCGGCGCGGCGGCGGCGTGTATGCGAC 840
Db 781 GCGGTGGGCGCAAGCGTCCGAACCCAGCGCGGCGGCGGCGGCGTGTATGCGAC 840
Qy 841 AACACCGCGCTTGGCGGTGCTCCGCAACGCGGCGAGATTTGTGCGGTCAACGACCTG 900
Db 841 AACACCGCGCTTGGCGGTGCTCCGCAACGCGGCGAGATTTGTGCGGTCAACGACCTG 900

```

```

Qy 901 CCGCGACAACCTGGGACCAAGATGCACTTCGAGATGATGCTGCCAACGTCATGTG 960
Db 901 CCGCGACAACCTGGGACCAAGATGCACTTCGAGATGATGCTGCCAACGTCATGTG 960
Qy 961 CAACCTCGGCGGAGATGAGGCTCTGCGGCCATCAACCCCAATCGCGGAGCGCGAG 1020
Db 961 CAACCTCGGCGGAGATGAGGCTCTGCGGCCATCAACCCCAATCGCGGAGCGCGAG 1020
Qy 1021 CGCGGTGTGCTGACGCGGTGCGCTTGCAGAACCGGCAAGATGTGCTTGTATGCGGCC 1080
Db 1021 CGCGGTGTGCTGACGCGGTGCGCTTGCAGAACCGGCAAGATGTGCTTGTATGCGGCC 1080
Qy 1081 TCGGCGCAATTCCTGTGCGGCGCGCTGATCACTACCAATATGATTCGCTGATGAC 1140
Db 1081 TCGGCGCAATTCCTGTGCGGCGCGCTGATCACTACCAATATGATTCGCTGATGAC 1140
Qy 1141 GAGACCGGCTTGTGACGCTGGAAGAGATGCGGTGTGAAAGAGCTGACGTTGATAT 1200
Db 1141 GAGACCGGCTTGTGACGCTGGAAGAGATGCGGTGTGAAAGAGCTGACGTTGATAT 1200
Qy 1201 GACGTGCGCGACCTTCTGCGGTGCGCGCGGAGCTGCTGACCGCACTGAACCGGAG 1260
Db 1201 GACGTGCGCGACCTTCTGCGGTGCGCGCGGAGCTGCTGACCGCACTGAACCGGAG 1260
Qy 1261 ACCGCAATTTACTTCTTGGCGCTGAACAATGCGCTAGATATTAAGCGCTGATCAA 1320
Db 1261 ACCGCAATTTACTTCTTGGCGCTGAACAATGCGCTAGATATTAAGCGCTGATCAA 1320
Qy 1321 GAGTTTACGCGCTGCACTGCTATTAACCAAGCGCGCAACCGCAACTCGCGCGGCTTT 1380
Db 1321 GAGTTTACGCGCTGCACTGCTATTAACCAAGCGCGCAACCGCAACTCGCGCGGCTTT 1380
Qy 1381 GAAATATGCGCGCGCATGCAAGCGATGATGATGACACCGGCGCACTTGTGTGCGG 1440
Db 1381 GAAATATGCGCGCGCATGCAAGCGATGATGATGACACCGGCGCACTTGTGTGCGG 1440
Qy 1441 GAGCGCGCTGCGGGAACCTACGCGCTTGTGCAAGCGGCGGTGTGATCAAC 1500
Db 1441 GAGCGCGCTGCGGGAACCTACGCGCTTGTGCAAGCGGCGGTGTGATCAAC 1500
Qy 1501 GCGCGGACCGACCGCTATTTTCGCTGCTGACGCAAGAAACCGCGAGACTTGTGTGCG 1560
Db 1501 GCGCGGACCGACCGCTATTTTCGCTGCTGACGCAAGAAACCGCGAGACTTGTGTGCG 1560
Qy 1561 GCCGCTCTGCGAGCGGTGCGGCAAGCGGCGGAGATGATGATGATGATGATGATG 1620
Db 1561 GCCGCTCTGCGAGCGGTGCGGCAAGCGGCGGAGATGATGATGATGATGATGATG 1620
Qy 1621 TATATGCGCATGCGGTGCGGCGGCTTGAACCTATGCGCAATTTGAACCGCGCGTGG 1680
Db 1621 TATATGCGCATGCGGTGCGGCGGCTTGAACCTATGCGCAATTTGAACCGCGCGTGG 1680
Qy 1681 GAGGCAATGATGATGACCTGCGGTGATGATGATGATGATGATGATGATGATGATG 1740
Db 1681 GAGGCAATGATGATGACCTGCGGTGATGATGATGATGATGATGATGATGATGATG 1740

```

RESULT 2

ADN10955 standard; DNA; 1740 BP.

```

XX ADN10955;
XX 01-JUL-2004 (first entry)
XX Gluconobacter oxydans Enzyme B coding sequence.
XX Enzyme B; ascorbic acid; vitamin C; L-gulono-1,4-lactone; L-gulononic acid;
XX L-galactono-1,4-lactone; L-galactonic acid; gene; ds.
XX Gluconobacter oxydans.
XX

```

Key	Location/Qualifiers
FT CDS	1..1740
FT	/*tag= a
FT	/product= "Enzyme B"
PN	MO2004029267-A1.
PD	08-APR-2004.
PP	22-SEP-2003; 2003MO-EP010489.
PR	27-SEP-2002; 2002EP-00021602.
PA	(STAM) DSM IP ASSETS BY.
PI	Hoshino T, Shinjoh M;
DR	WPI; 2004-329889/30.
XX	P-PSDB; ADN10956.
PT	Producing L-aescobolic acid using enzyme B of Gluconobacter oxydans, from
XX	substrates L-glucose, L-galactose, L-idose, and L-talose.
XX	
XX	Disclosure; SEQ ID NO 1; 24pp; English.
XX	
XX	The present sequence is the coding sequence for Enzyme B of Gluconobacter
XX	oxydans strain DSM 4025. Enzyme B has a molecular weight of about 60,000
XX	Da by SDS-PAGE, substrate specificity for primary and secondary alcohols
XX	and aldehydes, is stable in the pH range 6-9 with optimal activity at
XX	about pH 8.0, and is inhibited by Cu ²⁺ , Zn ²⁺ , Mn ²⁺ , Fe ²⁺ and Fe ³⁺ . The
XX	present invention provides the use of this enzyme in a process for
XX	producing L-aescobolic acid from L-glucose, L-galactose, L-idose or L-
XX	talose, or from L-gulono-1,4-lactone, L-gulonic acid, L-galactono-1,4-
XX	lactone, L-galactonic acid, L-idono-1,4-lactone, L-idonic acid, L-talono-
XX	1,4-lactone and L-talonic acid. Enzyme B is also used in a process for
XX	the production of L-gulono-1,4-lactone or L-gulonic acid from L-glucose,
XX	and L-galactono-1,4-lactone or L-galactonic acid from L-galactose. The
XX	processes involve contacting the enzyme with the respective substrate and
XX	isolating the product from the reaction mixture. The process is conducted
XX	for 1-70 hours at pH 1-9 (preferably pH 2-8) and 13-45 (preferably 18-42)
XX	degrees C. Production of L-gulono-1,4-lactone/L-gulonic acid from L-
XX	glucose, vitamin C from L-gulono-1,4-lactone/L-gulonic acid, L-galactono-
XX	1,4-lactone/L-galactonic acid from L-galactose, and vitamin C from L-
XX	galactono-1,4-lactone/L-galactonic acid by recombinant <i>Escherichia coli</i>
XX	UM109 carrying the Enzyme B gene is described in examples from the
XX	invention.
XX	
XX	Sequence 1740 BP; 340 A; 529 C; 545 G; 326 T; 0 U; 0 Other;
XX	
XX	Query Match 100.0%; Score 1740; DB 12; Length 1740;
XX	Best Local Similarity 100.0%; Pred. No. 0;
XX	Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY	1 ATGAACCCACAAACGCTGCTTGACCAAGCGCGCGCTGCTATTGCTTACCGCGCCGCC 60
DB	1 ATGAACCCACAAACGCTGCTTGACCAAGCGCGCGCTGCTATTGCTTACCGCGCCGCC 60
OY	61 GCATTGCGGCAAGTAAACCCCAATTACCGATGAACCTGCTGCGAACCAGCCGCCCTGTGAA 120
DB	61 GCATTGCGGCAAGTAAACCCCAATTACCGATGAACCTGCTGCGAACCAGCCGCCCTGTGAA 120
OY	121 TGGATTAACTACGCGCGCAACCAAAAACCTATCCCACTCGCCCTGTGACCAAGTACT 180
DB	121 TGGATTAACTACGCGCGCAACCAAAAACCTATCCCACTCGCCCTGTGACCAAGTACT 180
OY	181 GCCGCAACGTTGCTCAGTTGTCGAACTGGTCTGGCGCCCGCGGATGAGCGCGGCGCTA 240
DB	181 GCCGCAACGTTGCTCAGTTGTCGAACTGGTCTGGCGCCCGCGGATGAGCGCGGCGCTA 240
OY	241 CAGGTCACGCGCATATATCATATGATGCGCGTGAATGTTATCTGGCAAAACCCCGATATGATC 300
DB	241 CAGGTCACGCGCATATATCATATGATGCGCGTGAATGTTATCTGGCAAAACCCCGATATGATC 300

QY	301	CAGCGCGTGGATGCGCAAA	CAGGCGGACTGATCTCTGGAA	CACCGCGCCAACTGCCGCC	360
Db	301	CAGCGCGTGGATGCGCAAA	CAGGCGGACTGATCTCTGGAA	CACCGCGCCAACTGCCGCC	360
QY	361	GTCCGCA	CGCTAAACGCC	CCAAAGGCGACCG	420
Db	361	GTCCGCA	CGCTAAACGCC	CCAAAGGCGACCG	420
QY	421	CTCTATTTCA	GGCTATGGGCAACA	CACTCTATCGCGCTGGATAT	480
Db	421	CTCTATTTCA	GGCTATGGGCAACA	CACTCTATCGCGCTGGATAT	480
QY	481	GTATTCGATGTCGAC	GTGATCGGCGGAA	GCAGCTTGA	540
Db	481	GTATTCGATGTCGAC	GTGATCGGCGGAA	GCAGCTTGA	540
QY	541	ATTGTCCCA	TGCGCTCATCTGTGCGGGTTT	CCACTCTCCAAATAT	600
Db	541	ATTGTCCCA	TGCGCTCATCTGTGCGGGTTT	CCACTCTCCAAATAT	600
QY	601	TTTATCTCGG	GGGCGACGATTCGCGCAGCGG	TGAGAGCTGTGCGCAACA	660
Db	601	TTTATCTCGG	GGGCGACGATTCGCGCAGCGG	TGAGAGCTGTGCGCAACA	660
QY	661	CAGCCGCGCA	AGAGGCTGACGAGACTT	GGGCGCAATGTT	720
Db	661	CAGCCGCGCA	AGAGGCTGACGAGACTT	GGGCGCAATGTT	720
QY	721	GGCGTCTGGGGT	CGAGATCACTATGATCCCGTGA	CGAACTTGCTGCTATGAGT	780
Db	721	GGCGTCTGGGGT	CGAGATCACTATGATCCCGTGA	CGAACTTGCTGCTATGAGT	780
QY	781	GGCGTGGG	CCCGACGCTGCGAA	CCCGACGCGCA	840
Db	781	GGCGTGGG	CCCGACGCTGCGAA	CCCGACGCGCA	840
QY	841	AAACA	CCCGCTTTGGGGTGGCGCTCCCGCA	CCGGCGGAGATTTGCTGCGCTCA	900
Db	841	AAACA	CCCGCTTTGGGGTGGCGCTCCCGCA	CCGGCGGAGATTTGCTGCGCTCA	900
QY	901	CCGCGCGCA	CAACTGGGACCA	AGATGCAAGTGTGTGCTGCGCAAC	960
Db	901	CCGCGCGCA	CAACTGGGACCA	AGATGCAAGTGTGTGCTGCGCAAC	960
QY	961	CAACCTCTCGG	CCGAGATGAGAGGCTGTGCGCGCA	CTAACCCCAATGCGGAGCGGCGAG	1020
Db	961	CAACCTCTCGG	CCGAGATGAGAGGCTGTGCGCGCA	CTAACCCCAATGCGGAGCGGCGAG	1020
QY	1021	CGCGCTGTGCTG	AGCGGCTGCGCTTGCA	AGCCGGCACATGTGTGTTGATGCGCGC	1080
Db	1021	CGCGCTGTGCTG	AGCGGCTGCGCTTGCA	AGCCGGCACATGTGTGTTGATGCGCGC	1080
QY	1081	TCGGCGCA	ATTCTGTGGGCGCGCTGATACCA	CTACCAATATGATGCGCTCGATGAC	1140
Db	1081	TCGGCGCA	ATTCTGTGGGCGCGCTGATACCA	CTACCAATATGATGCGCTCGATGAC	1140
QY	1141	GAGAC	CCGGCCTGTGACGCTGAA	ACGAGAGTGGCGTGTGAAAGCTGCGACGTTGATAT	1200
Db	1141	GAGAC	CCGGCCTGTGACGCTGAA	ACGAGAGTGGCGTGTGAAAGCTGCGACGTTGATAT	1200
QY	1201	GACGCTGCGCGA	CTTCTGTGGTGGCGCGACTGTGCT	CAAGCCGCACTGAACCCGCGAC	1260
Db	1201	GACGCTGCGCGA	CTTCTGTGGTGGCGCGACTGTGCT	CAAGCCGCACTGAACCCGCGAC	1260
QY	1261	ACCGGCA	TTTATCTTCTTGTGCGCGCTGAA	CAAGTCTGCTGCAATATTA	1320
Db	1261	ACCGGCA	TTTATCTTCTTGTGCGCGCTGAA	CAAGTCTGCTGCAATATTA	1320
QY	1321	GAGTTT	AGCGCGCTCGACGCTTAA	CACACAGCGCGACCCGCAAA	1380
Db	1321	GAGTTT	AGCGCGCGCTCGACGCTTAA	CACACAGCGCGACCCGCAAA	1380
QY	1381	GAAAA	TATGGGCGCGCATCGACCGCA	TGATATACACCGGCGCACCTTGTGCTGCGG	1440


```

Db      1381  GAAATATATGGGCGCATGACGCGATGATATACAGACCGGCGCACCTTGTGTCGGCG 1440
QY      1441  GAGCGGCGCTGGCGGCAACTGCTGGCTTTGTGAGCGGCAAGCGGCTGGTTCAC 1500
Db      1441  GAGCGGCGCTGGCGGCAACTGCTGGCTTTGTGAGCGGCAAGCGGCTGGTTCAC 1500
QY      1501  GCGCGGACGCGACCGCTATTTCGCTGCGCTCAGCCAGAGAAACCGGCGAGACTTTGCGAG 1560
Db      1501  GCGCGGACGCGACCGCTATTTCGCTGCGCTCAGCCAGAGAAACCGGCGAGACTTTGCGAG 1560
QY      1561  GCGCGGCTTTCGCGACCGCTGCGAGCGGCGAGCGAGTACGATGAGTGGACCGCGTGA 1620
Db      1561  GCGCGGCTTTCGCGACCGCTGCGAGCGGCGAGCGAGTACGATGAGTGGACCGCGTGA 1620
QY      1621  TATATGCGCATCGCTGCGGCGGCTGACCTATGCGACCGCAATTGACCGCGCTGCGC 1680
Db      1621  TATATGCGCATCGCTGCGGCGGCTGACCTATGCGACCGCAATTGACCGCGCTGCGC 1680
QY      1681  GAGCGCATGATGCGACTGCGGCTGCGATGCGATGCGATGCGATGCGATGCGATGCG 1740
Db      1681  GAGCGCATGATGCGACTGCGGCTGCGATGCGATGCGATGCGATGCGATGCGATGCG 1740

```

RESULT 3

AD134117
ID AD134117 standard; DNA; 1743 BP.

AD134117;

15-APR-2004 (first entry)

Ketoglulonigenium sp. sorbitol dehydrogenase (SDH) 3 DNA.

Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulonic acid;
2KLG; gene; ds.

Ketoglulonigenium.

Location/Qualifiers

1..1743

/*tag= b

/product= "SDH protein"

1..69

/*tag= a

70..1740

/*tag= c

/product= "Mature SDH protein"

US2003228672-A1.

11-DEC-2003.

06-JUN-2002; 2002US-00162713.

06-JUN-2002; 2002US-00162713.

(ARCH) ARCHER-DANIELS MIDLAND CO.

Choi E, D'elia J, Kim H, Kim M, Lee UK, Pan J, Stoddard SF;

Yum D;

WPI; 2004-052025/05.

P-PSDB; AD134121.

New proteins of sorbitol dehydrogenases and cytochrome c of the strains
Ketoglulonigenium spp. useful in molecular biology, bacteriology and
industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
gulonic acid.

Claim 30; SEQ ID NO 4; 68pp; English.

CC The invention relates to the identification and isolation of nucleic acid
CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
CC of the strains, Ketoglulonigenium sp. The proteins and nucleic acid
CC molecules are useful in the fields of molecular biology, bacteriology and
CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-
CC gulonic acid (2KLG). The present sequence is Ketoglulonigenium sp. SDH
CC DNA.

XX Sequence 1743 BP; 361 A; 619 C; 496 G; 267 T; 0 U; 0 Other;

SO Query Match 69.3%; Score 1206.2; DB 12; Length 1743;

Best Local Similarity 80.9%; Pred. No. 1e-292;

Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;

```

QY      1  ATGAACCCCAAGCGCTGCTTCCGACAGCGGCGCGCTGCTATGCTTACCGGCGCCGC 60
Db      1  ATGCGACCCCAAGCGCTGCTTCCGACAGCGGCGCGCTGCTATGCTTACCGGCGCCGC 60
QY      61  GCATTGCGGAGGTAACCCCGATTAACCGATGACTCTGCGGCAACCCGCGCGTGTGA 120
Db      61  GCATTGCGGAGGTAACCCCGATTAACCGATGACTCTGCGGCAACCCGCGCGTGTGA 120
QY      121  TGGATTACTACGCGCGCAACCAAGAACTATGCGCACTGCGCCCTGACCCGATCACT 180
Db      121  TGGATTACTACGCGCGCAACCAAGAACTATGCGCACTGCGCCCTGACCCGATCACT 180
QY      181  GCGGACAAAGTGGTCACTGTCAGACTGCTGAGGCGCGGAGTGAAGCGGCGCGTA 240
Db      181  ACCGACAAAGTGGTCACTGTCAGACTGCTGAGGCGCGGAGTGAAGCGGCGCGTA 240
QY      241  CAGGTCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
Db      241  CAGGTCAAGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY      301  CAGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
Db      301  CAGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
QY      361  GTGCGCAGCTTAAACGCGCAAGCGCAACCGGCGCGCTGCGCTTTACCGGACAGC 420
Db      361  GTGCGCAGCTTAAACGCGCAAGCGCAACCGGCGCGCTGCGCTTTACCGGACAGC 420
QY      421  CTCATTTCAGCTCATGAGCAACCATCTGATGCGCGTGGATTTGAGAGCGGCGAGTC 480
Db      421  CTCATTTCAGCTCATGAGCAACCATCTGATGCGCGTGGATTTGAGAGCGGCGAGTC 480
QY      481  GTATTGATGTCGAAAGTGGATCGGCGGAGAGAGCGGCTTGAACCAAGCAACGAGG 540
Db      481  GTATTGATGTCGAAAGTGGATCGGCGGAGAGAGCGGCTTGAACCAAGCAACGAGG 540
QY      541  ATTGTGCCAATGCGCTCATGCTGCGGCTTCACTGCGCAATATGCGCTTATGATGC 600
Db      541  ATTGTGCCAATGCGCTCATGCTGCGGCTTCACTGCGCAATATGCGCTTATGATGC 600
QY      601  TTTATCTCGGGGCAAGTTCGCGGAGCGGCTGAGAGCTGAGGCAACCACTTTATCCG 660
Db      601  TTTATCTCGGGGCAAGTTCGCGGAGCGGCTGAGAGCTGAGGCAACCACTTTATCCG 660
QY      661  CAGCGGCGCAAGAGGCTGAGAGACTTGGGGCAATGATTTGAGGCGCGCTGATGAC 720
Db      661  CAGCGGCGCAAGAGGCTGAGAGACTTGGGGCAATGATTTGAGGCGCGCTGATGAC 720
QY      721  GCGCTTGGGGTCAAGTCACTATGATCCGCTACCAATGATGATGATGATGATGATG 780
Db      721  GCGCTTGGGGTCAAGTCACTATGATCCGCTACCAATGATGATGATGATGATGATG 780
QY      781  GCGCTGAGGCGGAGGCTGAGAGCAAGCGGCGGAGCGGCGGAGCGCTGTATGAC 840
Db      781  GCGCTGAGGCGGAGGCTGAGAGCAAGCGGCGGAGCGGCGGAGCGCTGTATGAC 840
QY      841  AACACCGGCTTTGCGGCTGCGGAGAGCGGCGGAGGAGTGTCTGCGGCTGACGAG 900
Db      841  AACACCGGCTTTGCGGCTGCGGAGAGCGGCGGAGGAGTGTCTGCGGCTGACGAG 900

```


QY 901 CCGGCGGACAACTGGGACCAAGATGACGTTGAGATGATGTCGCCCAACGTCGATG 960
 DB 901 CCGGCGGACAACTGGGACCAAGATGACGTTGAGATGATGTCGCCCAACGTCGATG 960
 QY 961 CAACCCCTCGGCGGAGATGAGAGGTGTCGCGGCATCAACCCCAATCGGCGAGCGAG 1020
 DB 961 CAGCGCGCTGCGGACATGAGCGGCGTTCAAGCCATCAACCCCAACCGCAACGCGAG 1020
 QY 1021 CCGCGGTGCTGACGGGTGCGCTTGGACAGACCGGACGATGATGTCGTTGATGCGGCC 1080
 DB 1021 CCGCGGTGCTGACGGGTGCGCTTGGACAGACCGGACGATGATGTCGTTGATGCGGCC 1080
 QY 1081 TCGGCGGAATTCCTGTTGGGCGCGGTGATACCACTAACCAATATGATCGGCTGATGAC 1140
 DB 1081 ACCGCGGAATTCCTGTTGGGCGCGGTGATACCACTAACCAATATGATCGGCTGATGAC 1140
 QY 1141 GAGACCGGCTTGTGACGATGAAAGAGATGCGATGCTGAAAGAGCTGAAATAT 1200
 DB 1141 GAGACCGGCTTGTGACGATGAAAGAGATGCGATGCTGAAAGAGCTGAAATAT 1200
 QY 1201 GAGGTGCGCGGACCTTCTGCGGTGAGGCGGCGATGCTGACCGGCGATGAAACCGGAG 1260
 DB 1201 GAGGTGCGCGGACCTTCTGCGGTGAGGCGGCGATGCTGACCGGCGATGAAACCGGAG 1260
 QY 1261 ACCGCGATTTACTTCTTGGCGGTGAAATGCTGCTGATGATATATGAGCGCTGATCAA 1320
 DB 1261 ACCGCGATTTACTTCTTGGCGGTGAAATGCTGCTGATGATATATGAGCGCTGATCAA 1320
 QY 1321 GAGTTAAGCGCGCTGACGCTTATACACAGCGGCGACCGCAAAATCTGCGCGGCTTT 1380
 DB 1321 GAGTTAAGCGCGCTGACGCTTATACACAGCGGCGACCGCAAAATCTGCGCGGCTTT 1380
 QY 1381 GAAATATGAGCGCGATGACGATGATGATGATGATGATGATGATGATGATGATGATG 1440
 DB 1381 GAAATATGAGCGCGATGACGATGATGATGATGATGATGATGATGATGATGATGATG 1440
 QY 1441 GAGCGCGCTTGGCGGAACTACTGCGCGCTTGTGCGAGCGGCGGTGATGATGATG 1500
 DB 1441 GAGCGCGCTTGGCGGAACTACTGCGCGCTTGTGCGAGCGGCGGTGATGATGATG 1500
 QY 1501 GCGCGGACCGGACCGCTATTTCCGCGGCGCTGACGCGGAAACCGGCGAGATTTGTGCGAG 1560
 DB 1501 GCGCGGACCGGACCGCTATTTCCGCGGCGCTGACGCGGAAACCGGCGAGATTTGTGCGAG 1560
 QY 1561 GCCCGTCTTGGCGAGCGGTGCGGACCGGCGAGCGGATGATGATGATGATGATGATG 1620
 DB 1561 ACCCGTCTGGCGAGTGTGCTTACCGGCGAGCGGATGATGATGATGATGATGATG 1620
 QY 1621 TATATGCGCATTCGTTGCGGCGGTCTGATCTTATGCGACCGAATTTGAAACGCGCGCTGGCG 1680
 DB 1621 TATATGCGCATTCGTTGCGGCGGTCTGATCTTATGCGACCGAATTTGAAACGCGCGCTGGCG 1680
 QY 1681 GAGGCAATCATTCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 1739
 DB 1681 GAGGCAATCATTCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT 1739
 RESULT 4
 ADI34124
 ID ADI34124 standard; DNA; 2281 BP.
 XX
 AC ADI34124;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Ketogulonigenium sp. sorbitol dehydrogenase (SDH) 3 gene.
 XX
 KM Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
 KM bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulononic acid;
 KM 2KLG; gene; ds.
 XX
 OS Ketogulonigenium.

XX
 PH Key Location/Qualifiers
 FT RBS 521..524
 FT
 FT CDS
 FT /tag= a
 FT /standard_name= "Shine-Dalgarno"
 FT 530..2272
 FT /tag= c
 FT /product= "SDH protein"
 FT 530..598
 FT /tag= b
 FT mat_peptide 599..2269
 FT /tag= d
 FT /product= "Mature SDH protein"
 PN US2003228672-A1.
 PD 11-DEC-2003.
 PD
 PF 06-JUN-2002; 2002US-00162713.
 PF
 PR 06-JUN-2002; 2002US-00162713.
 PR
 PA (ARCH) ARCHER-DANIELS MIDLAND CO.
 PI Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,
 PI Yum D;
 DR WPI: 2004-052025/05.
 DR P-PSDB; ADI34121.
 PS
 PS Claim 53; SEQ ID NO 11; 68bp; English.
 CC
 CC The invention relates to the identification and isolation of nucleic acid
 CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
 CC of the strains, Ketogulonigenium sp. The proteins and nucleic acid
 CC molecules are useful in the fields of molecular biology, bacteriology and
 CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-
 CC-gulononic acid (2KLG). The present sequence is Ketogulonigenium sp. SDH
 CC gene.
 XX
 SX Sequence 2281 BP; 483 A; 799 C; 632 G; 367 T; 0 U; 0 Other;
 Query Match 69.3%; Score 1206.2; DB 12; Length 2281;
 Best Local Similarity 80.9%; Pred. No. 1.1e-292;
 Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;
 QY 1 ATGAACCCCAACGCTGCTTGGACACAGCGCGCGGTGCTATTGCTTACCGCGCGCGC 60
 DB 530 ATGCACCCCAACACGCTGCTTGGACACAGCGCGCGGTGCTATTGCTTACCGCTCCT 589
 QY 61 GGATTGCGGCGAGTAAACCGGATTAACGATGAACTGCTGCGGAAACCGCGCGGTGTA 120
 DB 590 GCTTTGGCGAGGTAAACCGGATTAACGATGAACTGCTGCGGAAACCGCGCGGTGTA 649
 QY 121 TGGATTAACGAGCGCGCAACCAAGAACTATGCGCATGCGCGCTGATACCGCAATCACT 180
 DB 650 TGGATTAACGAGCGCGCAACCAAGAACTATGCGCATGCGCGCTGATACCGCAATCACT 709
 QY 181 GCCGCAACGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
 DB 710 ACCGCAACGTTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 769
 QY 241 CAGGTCAACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
 DB 770 CAGGTCAACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 829
 QY 301 CAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360

Db 830 CAGGCCATCGACGCCAAGACCGGGAACCTGATGAGGAAACCGCCGCACTGCCGCC 889
 Qy 361 GTGCCACGCTAAACGCCCAAGCGACCGCAACGCGCGCTGCCCTTTACGGCAGAGAC 420
 Db 890 GTTGCCCTCGGTGAACGCCCAAGCGACCGCTAAACGGCGCGTCCCTCTATGACCAAC 949
 Qy 421 CTGATTTAGCTCATGAGCAACAACCTGATGCGCTGGAATATGAGAGCGGCGAGTC 480
 Db 950 CTCTATTTACCTCGTGGCAACCACTTGTGCACTGGACATGGCCACCGCCCAAGTC 1009
 Qy 481 GTATTCGATGTCGACGTGATCGGCGGAAGACCGCTTTGACAGTAAACCAACGCGGCCG 540
 Db 1010 GTCTTGTGATGTCGACGCGCGCTCGGCGGATGACGCGCTGACAGCAACCAACGCGGCCG 1069
 Qy 541 ATTGTGCGCAATGCGCGCTCATGCTCGGCGGTTCACCGTGCAGATATTCGCGCTATGATGC 600
 Db 1070 ATTGTGCGCAACGCGCTCATGCTCGCGCTGCACTTGCAGATATTCGCGCTATGATGC 1129
 Qy 601 TTTATTCGCGGAGCGATTCGCGGACGCGGTGAGAGCTGTGGCGCAACCACTTTATCCCG 660
 Db 1130 TTTGTCTCGGGTCAAGCCCGCACGCGGGAAGAACTGTGGCGCAACTTCACTTCCCG 1189
 Qy 661 CAGCGCGCGAAGAGGTGACGAGACTTGGGCGAATGTTTGAAGCGCGCTGAGTACC 720
 Db 1190 CAGCGCGCGAAGAGGTGACGAGACTTGGGCGAAGTTCGATTCGATTCGCGCTGATGAC 1249
 Qy 721 GGGCTGTGGGTGATGATCACTATGATCCCGTACGACACTTGTGTTCTATGCGCTGACC 780
 Db 1250 GGGGTCTGGGCGACGATGATCTATGACCCCGTCAACCACTGATGACGCTGATGACC 1309
 Qy 781 GGGGTGGGCGGACGCTTCCAAACCCGCGGACCGCGCGCGCGCGCTGATGAGCACC 840
 Db 1310 GGGGTGGGCGGCGGCTTCCAAACCCGCGGACCGCGCGCGCGCGCTGATGAGCACC 1369
 Qy 841 AACACCGCGTTTCCGCTGCGTCCCGCACGCGGCGAATTTGCTGCGGTGACGACGCTGC 900
 Db 1370 AATACCGCGTTTCCGCTGCGTCCCGCACGCGGCGAATTTGCTGCGGTGACGACGCTGC 1429
 Qy 901 CCGCGCGACAACTGGGACCAAGATGCAAGTGTGAGATGATGCTGCCCACTGATGTCG 960
 Db 1430 CCGCGCGACAACTGGGACCAAGATGCAATTCGATGATGATGCTGCCCACTGATGTCG 1489
 Qy 961 CAACCCCTCGCGGAGATGAGGCGTGGCGCGCATCAACCCCAATCGCGGACGCGGACG 1020
 Db 1490 CAGCGCGCTGCGACATGACGCGCTTCAAGCGCATCAACCCCAATCGCGGACGCGGACG 1549
 Qy 1021 CCGCGGTGCTGACGCGGTGCGCTTCAAGACCGGACGATGCTGCTTATGATGCGGCC 1080
 Db 1550 CGTCCGCTGCTGACGCGGATCCCTGCAAAACCGGACGATGCGATGCGACGCGGAC 1609
 Qy 1081 TCGGCGGAATTCCTGTTGGGCGCGTGAATCACTACCAATATGATGCGCTCGATGAC 1140
 Db 1610 ACCGCGCAATTCCTGTTGGGCGCGTGAATCACTACCAATATGATGCGCTCGATGAC 1669
 Qy 1141 GAGACCGGCGCTTGTGACGCTGAACGAGATGCGGTGCTGAAGAGCTGACGCTGATAT 1200
 Db 1670 GAAACCGGCGCTTGTGACGCTGAACGAGATGCGGTGCTGAAGAGCTGACGCTGATAT 1729
 Qy 1201 GAGCTCTGCGCGACCTTCTCGGTGGGCGGATGCTGCTGACGCGGCACTGAACCGGAC 1260
 Db 1730 GACATCTGCGCGACCTTCTCGGTGGGCGGATGCTGCTGACGCGGCACTGAACCGGAC 1789
 Qy 1261 ACCGCGATTTACTTCTTGGCGCTGAACCAATGCTGCTGAATATTTGACGCTGATCA 1320
 Db 1790 AGGGGATCTACTTCAATACCGCTGAACCAACGCTGCTGATCACTATGCGCTGATCA 1849
 Qy 1321 GAGTTTAGCGCGCTCGACGCTTATTAACCAAGCGGACCGCAAAATCGCGCGCGCTTT 1380
 Db 1850 GAATTTCTCGCGCTTATGATGTAACAATACGAGGCACTCTTAACAGTTTGCACCGGCGCTTT 1909
 Qy 1381 GAAATATATGCGCGGATTCGACGCGATGATATATGAGACCGGCGGACCTTTGATGCGGCG 1440
 Db 1910 GAAAAATATGCGCGGATTCGACGCGATGATATATGAGACCGGCGGAAACCTTGTGTGCGGCG 1969

Qy 1441 GAGGCGCCCTGCGGGAACCTACTGCGCGCTTTTGTGACGCGGACGCGGTGTGTTCAAC 1500
 Db 1970 GAACTCTGCGGTGAACCTACTGCGCGCTTCTCTGACGCGCTGCGCGGTGTGTTCAAC 2029
 Qy 1501 GCGCGGAACCGACCGCTATTTCCGTGCGCTCAGACGAGAAACCGGACGACTTTGTGGCAG 1560
 Db 2030 GCGCGCAACCGACCGGTACTTGTGCGCGCTCAGACGAGAAACCGGACGACTTTGTGGCAG 2089
 Qy 1561 GCGCGCTTGTGACGCGTCCGACGCGGCGAGGACATCAAGATGATGAGCGGCGTGGCAA 1620
 Db 2090 ACCGCTCTGCGGATGTCGCTACCGGCGCAAGCACTACGAAATCGACGGGACCTAA 2149
 Qy 1621 TATATCCCATGCGGTGCGGCGGTCTGACCTATGACGCAAGATGAAACGCGCGCTGGCC 1680
 Db 2150 TACGTCCGATTCGCGGGGGGGGACGACCTACGCGCAACCAACCAACCGTCCCTAGC 2209
 Qy 1681 GAGGCATTCGATTTGACCTCGGTGCTGATTAAGCATCTATGCTTTGACCTGCGGAGTA 1739
 Db 2210 GAGGCGATCGACTGACCAAGATCGGCAACGCGGTTTACGTTTGTGCGTGGCGACGA 2268

RESULT 5

ID AAV29051 standard; CDNA; 1740 BP.
 AAV29051;

21-AUG-1998 (first entry)

Alcohol and/or aldehyde dehydrogenase A nucleotide sequence.

KM Alcohol/aldehyde dehydrogenase A enzyme; recombinant organism; aldehyde;
 KM ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulononic acid;
 KM L-ascorbic; inhibition; ds.

Glucobacter oxydans.

FH Key Location/Qualifiers
 FT 1..1737
 FT /*tag= a
 FT /product= "alcohol and/or aldehyde dehydrogenase A"

EP832974-A2.

01-APR-1998.

11-SEP-1997; 97EP-00115801.

19-SEP-1996; 96EP-00115001.

(HOPE) HOFFMANN LA ROCHE & CO AG F.

Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;

WPI; 1998-195228/18.

P-PSDB; AAW37873.

PT Recombinant Glucobacter oxydans alcohol and/or aldehyde dehydrogenase
 PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-
 PT sorbitol to 2-keto-L-gulononic acid.

PS Disclosure; Page 27-28; 59pp; English.

This is the nucleotide sequence encoding the Glucobacter oxydans
 alcohol and/or aldehyde dehydrogenase A enzyme. The enzymes or
 recombinant organisms can be used to convert suitable substrates to
 aldehydes, ketones or carboxylic acids, especially to convert L-sorbose
 or D-sorbitol to 2-keto-L-gulononic acid, which can be converted to L-
 ascorbic acid by standard procedures. The derivatives of ADH enzymes
 have desired substrate specificity, higher affinity to a substrate, lower
 affinity to an inhibitory compound, higher stability against temperature
 and/or pH and higher catalytic speed

XX Sequence 1740 BP; 356 A; 558 C; 507 G; 319 T; 0 U; 0 Other;
 SQ Query Match 65.0%; Score 1130.2; DB 2; Length 1740;
 Best Local Similarity 78.8%; Pred. No. 1.3e-273;
 Matches 1374; Conservative 0; Mismatches 363; Indels 6; Gaps 2;

QY 1 ATGAACCCCAACAGCTGCTTCCACCAGCGCGCGTGTATGCTTACCGCGCCGCC 60
 DB 1 ATGAACCGACTCGGCTTGGGCGACGTGGGCACTTGCACTTGCTGCGGACCC 60

QY 61 GCATTCGGGAGGTAACCCCGATTACCGATGAACCTGTGCGAACCCCGCGCTGTGA 120
 DB 61 GCCTTGTCTAAATGACCCCGTACCGATGATTTGTGCGAACCCCGCGCTGTGA 120

QY 121 TGGATTACTACGCGCGCAACCAAGAAATCTATCGCACTCGCCCTGACCCAGATCACT 180
 DB 121 TGGATCAGCTACCGTGAACCAAGAAATCTATCGCACTCGCCCTGACCGATCACT 180

QY 181 GCCGACAAGCTGTGTAATGTTGCACTGTGTGCGCGCGGAGTGAAGCGCGGCGCTA 240
 DB 181 ACTGAGACGCTCGGCAACTGCAACTGTGTGCGCGCGGCAATGACCGCGCAAGTC 240

QY 241 CAGGTACCGCGAGTATCCATGATGCGGTGATGATCTGCGCAACCCCGGTATGATGATC 300
 DB 241 CAGGTACCGCGCTGATCCATGATGCGGTGATGATCTGCGCAACCCCGGTATGATGATC 300

QY 301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGAACACCCCGCGCACTGCGCGC 360
 DB 301 CAGGCGCTGACGCAAACTGGCGATCTGATCTGGAACACCCCGCGCACTGCGCGAC 360

QY 361 GTGCGCACGTTAAAGCCGCAAGCGGACCGCAAGCGCGCGTCCGCTTTTACGCAAGAC 420
 DB 361 ATGCGCACGTTGAACGCTTTGCGAGCGCACCGCGCATGCGGTGAACGCGCAACAC 420

QY 421 CTCTATTTCAGCTCATGAGCAACCACTCTGATCGCGCTGATATGAGACGCGCGCATGTC 480
 DB 421 GTTACTCTTGTTCGTGAGCAACCACTCTGATCGCGCTGATATGAGACGCGCGCATGTC 480

QY 481 GTATTGATGTGAAAGCTGTGATCGGCGAAAGACGCGCTTGAACAGTAAACACACGCGGCGC 540
 DB 481 ACCTTGCACGCTGACCGCGCGCAAGGCGAAGACATGATGCTTCG---AACTGTCGCGCGCGC 537

QY 541 ATTGTGCGCAATGCGGCTCATGCTGCGGCGGTTCACCTGCGCAATTCGCGCTATGATGATG 600
 DB 538 ATGTGCGCAACGCGCTGATGCTGCGGCTTCGATGCACTGCGCAATTCGCGCTATGATGATG 597

QY 601 TTTATCTCGGCGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCACTTATCCCG 660
 DB 598 TTTGTCTCGGCGGACGATTCGCGGACGCGGTGAGAGCTGTGCGCAACCACTTATCCCG 657

QY 661 CAGCCGCGGCAAGAGGTGACGAGCTTGGGCGAAATGATTTGAGGCGCGCTGTGATGAGAC 720
 DB 658 CGCGCTGCGCAAGAGGTGATGAGCTTGGGCGAAAGATTTGAGGCGCGCTGTGATGAGAC 717

QY 721 GCGCTGCGGCTGATCACTATGATCCCGTACGAACTTGTGTTCTATGCTGACCC 780
 DB 718 GGTGCTGCGGCGGATCACTATGATCCCGTACGAACTTGTGTTCTATGCTGACCC 777

QY 781 GCGGTGCGGCGGATCCGATCCGAAACCCAGCGGCGGACCGCGCGGCGCTGTATGAGAC 840
 DB 778 GGTGTGCGGCTCGGCTCGGAAACCCAGCGGCGGACCGCGCGGCGCTGTATGAGAC 837

QY 841 AACACCGGCTTTCGCGGTGCTCCGCAACGCGCGAGATTTGTTGCGGTACCGCAACCTG 900
 DB 838 AACACCGGCTTTCGCGGTGCTCCGCAACGCGCGAGATTTGTTGCGGTACCGCAACCTG 897

QY 901 CCGCGCGACAACTGCGGACCAAGATGCACTTGAAGATGATGCTGCGCAACGTCGATGTG 960
 DB 898 CCGCGCGACAACTGCGGACCAAGATGCACTTGAAGATGATGCTGCGCAACGTCGATGTG 957

QY 961 CAACCCCTCGGCGGAGATGAGAGGTGCGCGCATCAACCCCAATGCGCGGACGCGCGAG 1020

DB 958 CAACCCCTGACGAGATGAGAAAGTCTGCACTGATCAACCCGAAACGCCCAACTGCGGAG 1017
 QY 1021 CGCGGTGCTGAGCGGAGTCCGCTTTGCAAGACCGGCAACGATGATGCTTGTGATGCGGCC 1080
 DB 1018 GGTGCGGTGCTGAGCGGAGTCCGCTTTGCAAGACCGGCAACGATGATGCTTGTGATGCGGCC 1077

QY 1081 TCGGCGGAATTCCTGTGCGGCGCGTATGCAACTACCAACCAATGATGCTGCGATGAC 1140
 DB 1078 ACCGCGAATTCCTGTGCGGCGCGTATGCAACTACCAACCAATGATGATGATGATGAC 1137

QY 1141 GAGACCGGCTTGTGACGCTGATGAGAGATGCGGTGCTGAAAGCTGTGAAATAT 1200
 DB 1138 GAAACCGGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197

QY 1201 GACGTCTCCCGACCTTCTCGGCTGCGGCGGACGATGCTGATGCGGCACTGAACCGGAC 1260
 DB 1198 GACGTCTCCCGACCTTCTCGGCTGCGGCGGACGATGCTGATGCGGCACTGAACCGGAC 1257

QY 1261 ACCGCGATTTACTTCTTCCGCTGAAACATGCTGCTACGATATTTATGCGCTGATCA 1320
 DB 1258 AGCGGCACTTACTTCTTCCGCTGAAACATGCTGCTACGATATGATGCGGCGGATCA 1317

QY 1321 GAGTTTACGCGCTGACGCTATTAACACGAGCGGACCGGAAACTGCGCGGCGCTTT 1380
 DB 1318 GATTTCACTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377

QY 1381 GAAATATGAGCGCGCATGACGCGGATGATATACACCGGCGGACCTTGTGCGCG 1440
 DB 1378 GATATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1437

QY 1441 GAGCGCGCTGCGGCGGATATCTCCGCTTTTGTGACGAGCGGAGCGGTGATGATGAC 1500
 DB 1438 GAACTGCTGCGGCGGATATCTCCGCTTTTGTGACGAGCGGAGCGGTGATGATGAC 1497

QY 1501 GCGCGGACCGACCGCTATTTCCGTCGCTCCTACGCGGAAACCGGCGGACCTTTGTGCGAG 1560
 DB 1498 GGTGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1557

QY 1561 GCGCGCTTGTGCGAGGTCGCAACGCGGCGGAGCATGACCTACGATGATGATGATGATGAT 1620
 DB 1558 ACCGCGCTTGTGCAACGCTGCGTCCGCGCGGCGGCACTCTTACGAGGTTGAGCGGATCA 1617

QY 1621 TATATGCGCATGCGGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1679
 DB 1618 TATGTCGCAATGCGAGGTCGATGATGATGATGATGATGATGATGATGATGATGAT 1677

QY 1680 --CGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737
 DB 1678 GCGCGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737

QY 1738 TPA 1740
 DB 1738 TPA 1740

RESULT 6
 AA29053
 ID AA29053 standard; cDNA; 1737 BP.

XX AA29053;
 DT 10-AUG-1998 (first entry)
 XX
 DE Alcohol and/or aldehyde dehydrogenase A', nucleotide sequence.
 XX
 KM Alcohol/aldehyde dehydrogenase A', enzyme; recombinant organism;
 KM aldehyde; ketone; carboxylic acid; L-sorbose; D-sorbitol;
 KM 2-keto-L-gulonidic acid; L-ascorbic; inhibition; ds.
 OS Gluconobacter oxydans.
 XX
 XX
 Key Location/Qualifiers
 FT CDS 1..1737

```

FT      /*tag= a
PN      /product= "alcohol and/or aldehyde dehydrogenase A'"
XX      EP832974-A2.
XX      01-APR-1998.
XX      11-SEP-1997; 97EP-00115801.
XX      19-SEP-1996; 96EP-00115001.
XX      (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX      Aakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N;
XX      WPI, 1996-195228/18.
XX      P-Psdb; AAm37875.
PT      Recombinant Glucobacter oxydans alcohol and/or aldehyde dehydrogenase
PT      enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-
XX      sorbitol to 2-keto-L-gulonic acid.
PS      Disclosure; Page 31-32; 59pp; English.
XX      This is the nucleotide sequence encoding the Glucobacter oxydans
XX      alcohol and/or aldehyde dehydrogenase A' enzyme. The enzymes or
XX      recombinant organisms can be used to convert suitable substrates to
XX      aldehydes, ketones or carboxylic acids, especially to convert L-sorbose
XX      or D-sorbitol to 2-keto-L-gulonic acid, which can be converted to L-
XX      ascorbic acid by standard procedures. The derivatives of ADH enzymes
XX      have desired substrate specificity, higher affinity to a substrate, lower
XX      affinity to an inhibitory compound, higher stability against temperature
XX      and/or pH and higher catalytic speed
XX      Sequence 1737 BP; 339 A; 570 C; 509 G; 319 T; 0 U; 0 Other;
Query Match 62.1%; Score 1079.8; DB 2; Length 1737;
Best Local Similarity 77.1%; Pred. No. 5,86-261;
Matches 1341; Conservative 0; Mismatches 392; Indels 6; Gaps 2;
QY      1 ATGACCCCAACAGCGTGTGTCAGCAGGCGCGCGTCTATGCTTACCGCGCCGC 60
DB      1 ATGAAACTGACGACCTGCTGCAAGACGCGCGCTGTGCTTGGACACATCCC 60
QY      61 GCATTGCGCGAGTAACCCCGATTACGATGAATCTGCTGCGAACCCTGCTGTGA 120
DB      61 GCCCTGCGCA---AACCGCATCACCGATGAATCTGCGAACCCTGCTGTGA 117
QY      121 TGGATTAACTACGCGCGCAACCAAGAACTATCGCACCTGCGCCCTGACCGAGTACT 180
DB      118 TGGATCAACTACGCTGCAAGACCAAGAACTACCGCACTGCGCCTGACGAGTACC 177
QY      181 GCCGCAACCGTTGTCAGTTGCAACTGATCTGCGGCCCGCGGATGAGAGCGGAGCCGTA 240
DB      178 GCAGACAACTGCGGCACTGCAACTGATCTGCGGCCCGCGGATGAGAGCGGAGATC 237
QY      241 CAGGTACCGCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
DB      238 CAGGTACCGCCGTTGTCAGTGAAGCGATGATGATGATGATGATGATGATGATGATG 297
QY      301 CAGGCGCTGATGCGCAACAGCGCATGATCTGAGGAAACCGCGCGCACTGCGCGCC 360
DB      298 CAGGCGCATGACCGCGCGACCGCGCATGATCTGAGGAAACCGCGCGCACTGCGCGA 357
QY      361 GTGCGCACTGTAACCGCGCAAGCGCGACCGCGCGCTGCGCTTTTACGCGACGAC 420
DB      358 ATGCGCACTGTAACCGCTTGTGATGAGCGCGACCGCGCGATGCGCTTATGCGACA 417
QY      421 CTGATTTTACGCTCATGAGCAACCACTGATGCGCTGATGATGAGAGCGGCGCGAGTC 480
DB      418 GTCTATTTTGTCTGTGAGCAACCACTGATGCGCTGATGAGAGCGGCGCGAGTC 477
QY      481 GTATTGATGTGACAGCTGATGCGCGCAAGACGCGCTTGAACGATTAACCAACGCGG 540

```

```

DB      478 GTATTGACGTCGATTCGCGGTCAAGGACCGAAT---TGTCTGCAACTGTGCGGCGG 534
QY      541 ATTGTGCCAATGCGCTGATGCTGCGGTTTCACTGCGCAATTTGCGCTTATGATGC 600
DB      535 ATTGTGCCAATGCGCTGATGCTGCGGCTTCACTGCTGCAATTTGCGCTTATGATGC 594
QY      601 TTTATCTCGGGGACGATTCGCGGAGCGGTGAGAGGTGTGAGGCAACCTTATCCG 660
DB      595 TTGCTTTGAGGCAACGATCTGCGCACCGGTGAAGAGCTGTGAGGCAACCTTATCCG 654
QY      661 CAGCGGCGCAAGAGGATGACGAGACTTGGGGCAATGATTTGAGGCGGCTGATGACC 720
DB      655 CGCGCGCGCAAGAGGATGATGAGACTTGGGGCAATGATTTGAGGCGGCTGATGACC 714
QY      721 GCGCTTGGGGTCAATCACTGATGATTCGCGTGAACGAACTTGTGATGAGCTGAC 780
DB      715 GCGCTTGGGGTCAATCACTGATGATTCGCGTGAACGAACTTGTGATGAGCTGAC 774
QY      781 GCGGTGAGCGCGAGCTGCGAAGCCAGCGCGGACCGCGCGCACTGATGAGCACC 840
DB      775 GCAATTGCGCGTGTGCGCGGATTCAGCGCGGACCGGCTGATGATGAGCACC 834
QY      841 AACCGCGCTTTGCGGTGCGTCCCGACAGCGGAGATTTGTGGGTGACCGACCGCC 900
DB      835 AACCGCGCTTTGCTGTGCGCGCGCGGACCGGAGATTCGTGGGTGACCGAACTCTG 894
QY      901 CCGCGCACTTGGGACCAAGATGACCTTGGAGTATGATGATGATGATGATGATGATG 960
DB      895 CCGCGCACTTGGGACCAAGATGACCTTGGAGTATGATGATGATGATGATGATGATG 954
QY      961 CAACCTTGGCGGAGATGAGAGGTCTGCGCGCATCAACCCAAATGCGGCGAGGCGAG 1020
DB      955 CAGCGCTGCGGTGAGTGAAGAGCGCTGACGCGATCAACCCGATCGCGCGAGGCGAG 1014
QY      1021 CGCGGTGCGTGAAGCGGTGCGCGCTTGGCAAGCGGACGATGATGATGATGATGATG 1080
DB      1015 GTGTGCTTTGACCGCGCGCTTCCGTGCAAGACGACCACTGTGCGATTCGACGCGAA 1074
QY      1081 TGGGCGCAATTCCTGTGAGGCGCGGTGATTAACAATCACTAATGATGATGATGAT 1140
DB      1075 ACCGCGCAATTCCTGTGAGGCGCGCGACACACGATTAAGAACTGATGAAAGCGT 1134
QY      1141 GAGACCGCGCTTGTGAGGCGCGGTGATTAACAAGCGGTGATGATGATGATGATG 1200
DB      1135 CCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1194
QY      1201 GACGTGCGCGACCTTCTGTGAGGCGCGGATGATGATGATGATGATGATGATGATG 1260
DB      1195 GAATGCGCGGACCTTCTGTGAGGCGCGGATGATGATGATGATGATGATGATGATG 1254
QY      1261 ACCGCGATTTACTTCTTGTGCGGCGGATGATGATGATGATGATGATGATGATGATG 1320
DB      1255 ACTGCGATTTACTTCTTGTGCGGCGGATGATGATGATGATGATGATGATGATGATG 1314
QY      1321 GAGTTAGCGGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
DB      1315 GAGTTAGCGGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1374
QY      1381 GAAATATGCGCGCGATGACGCGATGATGATGATGATGATGATGATGATGATGATG 1440
DB      1375 GAAATATGCGCGCGATGACGCGATGATGATGATGATGATGATGATGATGATGATG 1434
QY      1441 GAGGCGCGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
DB      1435 GAGGCGCGTGTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1494
QY      1501 GCGCGGACCGACCGCTATTTTCTGTGCGCGGATGATGATGATGATGATGATGATG 1560
DB      1495 GCGCGGACCGACCGCTATTTTCTGTGCGCGGATGATGATGATGATGATGATGATG 1554
QY      1561 GCGCGTCTTGTGAGCGGTGAGGCGGATGATGATGATGATGATGATGATGATGATG 1620

```

Db 1555 ACCGCTGCGCATGCTGCGCTCGGCTCAAGAGATTTCTATGATGACGCGCTGCA 1614
QY 1621 TATATGCGCATCGGTGCGGCGGTCTGACTATGCGACGAAATTGAACGCGCGCTGGCC 1680
Db 1615 TATGTCGCATCGGCGCGGCGGACACAGTATGCGACCAACACGCGCGCTGACG 1674
QY 1681 GAGGCAATGATTCGACCTGCGGTAAATGCGATCTATGCTTTGACCTGCGCGAGTA 1739
Db 1675 GAGCGATGACATCGACCGCATCGGACGCGCATCTATGTCTTCTGCTGCGCGAGCA 1733
RESULT 7
AD134114
ID AD134114 standard; DNA; 1737 BP.
XX AD134114;
XX
XX
XX 15-APR-2004 (first entry)
XX
XX Ketogulonigenium sp. sorbitol dehydrogenase (SDH) 1 DNA.
XX
XX Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
XX bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulonic acid;
XX 2KLG; gene; ds.
XX
XX Ketogulonigenium.
XX
XX
XX Key Location/Qualifiers
XX CDS 1..1737
XX sig_peptide /*tag= b
XX /*product= "SDH protein"
XX mat_peptide /*tag= a
XX 70..1734
XX /*tag= c
XX /*product= "Mature SDH protein"
XX
XX US2003228672-A1.
XX
XX 11-DEC-2003.
XX
XX 06-JUN-2002; 2002US-00162713.
XX
XX 06-JUN-2002; 2002US-00162713.
XX
XX (ARCH) ARCHER-DANTELS MIDLAND CO.
XX
XX Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF;
XX Yum D;
XX
XX WPI; 2004-052025/05.
XX P-PSDB; AD134118.
XX
XX New proteins of sorbitol dehydrogenases and cytochrome c of the strains
XX PT Ketogulonigenium spp., useful in molecular biology, bacteriology and
XX PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
XX PT gulonic acid.
XX
XX Claim 3; SEQ ID NO 1; 68pp; English.
XX
XX The invention relates to the identification and isolation of nucleic acid
XX CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
XX CC of the strains, Ketogulonigenium sp. The proteins and nucleic acid
XX CC molecules are useful in the fields of molecular biology, bacteriology and
XX CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-
XX CC gulonic acid (2KLG). The present sequence is Ketogulonigenium sp. SDH
XX CC DNA.
XX
XX Sequence 1737 BF; 366 A; 556 C; 496 G; 319 T; 0 U; 0 Other;
XX
XX
XX Query Match 61.6%; Score 1071.2; DB 12; Length 1737;
XX Best Local Similarity 76.4%; Pred. No. 8.4e-259;
XX Matches 1329; Conservative 0; Mismatches 408; Indels 3; Gaps 1;

QY 1 ATGAACCCCAACAAGCTGCTTGGACGACGCGCGCGTATGCTTACCGCGCGCC 60
Db 1 ATGAATGAAATGCTGCTTCTTGGCAAGCTTCTGCTTGCAATCTTTGCTGTC 60
QY 61 GCATTGCGCAGGTAAACCCCGATTAACGATGAACTGTGCGAAACCGCGCGTGTAA 120
Db 61 GCATTGCGCAGGTAAACCCCGATTAACGATGAACTGTGCGAAACCGCGCGTGTAA 120
QY 121 TGAATTAATAAGCGCGGCAACCAAGAAATCTATGCGCACTGCGCCCTGACCAATCACT 180
Db 121 TGAATTAATAAGCGCGGCAACCAAGAAATCTATGCGCACTGCGCCCTGACCAATCACT 180
QY 181 GCGGCAAGTGTGCTCACTGTTCACTGCTGCGCGCGCGGATGAGAGCGCGCGCTA 240
Db 181 GCGGCAAGTGTGCTCACTGTTCACTGCTGCGCGCGCGGATGAGAGCGCGCGCTA 240
QY 241 CAGGTACAGCGCGATGATTCATGATGCGGTGATGTATCTGCGAAACCCCGGTGATGATC 300
Db 241 CAGGTACAGCGCGATGATTCATGATGCGGTGATGTATCTGCGAAACCCCGGTGATGATC 300
QY 301 CAGGCGCTGAGTGGCAAAACAGGAGATCTGATGAGAAACCGCGCGCAACCTGCGCGC 360
Db 301 CAGGCGCTGAGTGGCAAAACAGGAGATCTGATGAGAAACCGCGCGCAACCTGCGCGC 360
QY 361 GTGCGCAAGCTPAAACGCCCAAGCGACCGCAAGCGCGCGCTGCGCTTACGCGACAGC 420
Db 361 ACCTGACGCTCAGCTCGCTGGGGATGCAAGCGCGGATCGCGCTTATGCGACCAAT 420
QY 421 CTCTATTCAAGTCAATGAGGAAACCATCTGATGCGCGTGAATGAGAACGCGCGCAAGTC 480
Db 421 CTCTATTCAAGTCAATGAGGAAACCATCTGATGCGCGTGAATGAGAACGCGCGCAAGTC 480
QY 481 GTATTGATGTGCAAGCTGATTCGCGGCAAGACCGGCTGACCAAGTAAACCAACCGCGCG 540
Db 481 GTATTGATGTGCAAGCTGATTCGCGGCAAGACCGGCTGACCAAGTAAACCAACCGCGCG 540
QY 541 ATTGTCGCAATGCGCTCATGCTGCGCGGTTCCACTGCGCAATTTGGCCCTATGATGC 600
Db 541 ATTGTCGCAATGCGCTCATGCTGCGCGGTTCCACTGCGCAATTTGGCCCTATGATGC 600
QY 598 TTGTCGCGGCGCATGATGCAAGACGCGCGCAAAACCTGCGCGCACTTCACTTCACTCCG 657
Db 598 TTGTCGCGGCGCATGATGCAAGACGCGCGCAAAACCTGCGCGCACTTCACTTCACTCCG 657
QY 661 CAGCGCGGCAAGAGGTGACGAACCTTGGGCAATGATTTGAGGCGCGCTGATGAC 720
Db 661 CAGCGCGGCAAGAGGTGACGAACCTTGGGCAATGATTTGAGGCGCGCTGATGAC 720
QY 721 GCGCTGCGGCGTCAAGATCACTATGATCCGTCGACGAACCTTGTGTTATGCTGAC 780
Db 721 GCGCTGCGGCGTCAAGATCACTATGATCCGTCGACGAACCTTGTGTTATGCTGAC 780
QY 778 GCGCTGCGGCGTCAAGATCACTATGATCCGTCGACGAACCTTGTGTTATGCTGAC 777
Db 778 GCGCTGCGGCGTCAAGATCACTATGATCCGTCGACGAACCTTGTGTTATGCTGAC 777
QY 841 AACACCGCTTTGGGTCGCTGCTGCGGCAAGATGCTGCGGTCACCAACCTG 900
Db 841 AACACCGCTTTGGGTCGCTGCTGCGGCAAGATGCTGCGGTCACCAACCTG 900
QY 897 AACACCGCTTTGGGTCGCTGCTGCGGCAAGATGCTGCGGTCACCAACCTG 897
Db 897 AACACCGCTTTGGGTCGCTGCTGCGGCAAGATGCTGCGGTCACCAACCTG 897
QY 901 CCGCGCACTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 960
Db 901 CCGCGCACTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 960
QY 957 CCGCGCACTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 957
Db 957 CCGCGCACTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 957
QY 961 CAACCTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 1020
Db 961 CAACCTGCGGCAAGATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 1020
QY 958 CAGCGCGTGGCGCAATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 1017
Db 958 CAGCGCGTGGCGCAATGCAAGTTCAGATGATGCTGCGCAACCTGATGATG 1017
QY 1021 CCGCGTGGTGAAGGCGCTTGGCAAGCGCGCAAGTGGTGGTGGTGGTGGTGG 1080
Db 1021 CCGCGTGGTGAAGGCGCTTGGCAAGCGCGCAAGTGGTGGTGGTGGTGGTGG 1080
QY 1018 CCGCGTGGTGAAGGCGCTTGGCAAGCGCGCAAGTGGTGGTGGTGGTGGTGG 1077
Db 1018 CCGCGTGGTGAAGGCGCTTGGCAAGCGCGCAAGTGGTGGTGGTGGTGGTGG 1077

OY	1081	TCGGGCGAAATTCCTGTGGGCGCGGATATCCAACTAACCAATATGATGGCTGCATGCAC	1140
Db	1078	ACGGGCGAAATTCCTGTGGGCGCGGACACCAACTACCAAAACATGATCAGTTCGATGCAC	1137
OY	1141	GAGACCGGCGCTTGTGACCGGTGACGAGGATCGGCTGCTGAAGAAGCTTGGACCTTGAATAT	1200
Db	1138	GAAACCGGCTGTGTGACCGGTGAATGAATATCATCTCAAAAAGATCTTGGACACCGCACTAC	1197
OY	1201	GACGCTTCCCGACCTTCTGGTGGGCGCGACACTGATGTACAGCGGCACTGAACCCGAGAC	1260
Db	1198	CGCATTTGCCCGACATCTTGTGGTGGAGCGGACGTGGCGGTGGGATCCTTGAACCCGGAT	1257
OY	1261	ACCGGCATTTACTTCTTGGCGGCTGAACAAATGCCCTGCTACGATATATATGCGCGTTGATCA	1320
Db	1258	AGCGGCATCTATCTTCAATTCCTCCCTGAACAAGCCTGTGTGGATTTGGCGGACGTGATCPA	1317
OY	1321	GAGTTTAGCGGCTCGACGTCTATTAACAACAGCGCGGACCGGCAAACTTCGCGCGGCGCTTT	1380
Db	1318	GAGTTCACGGCAATGAGAGGTCTTACACACAGCGCGGCACTTCTCGCTTGGCGCGCGGAAAA	1377
OY	1381	GAAATATATGGGCGCGCATGAGCGGCAATGATATATACGACCCGGGCGCACCTTGTGGTGGCG	1440
Db	1378	GAAATCATGGCGCGCATGAGCGGCAATGACATCAAGACGGGCAAAACCTGTGTGTGGCTC	1437
OY	1441	GAGCGCCCTGCGGCGCACTACTCGCCGCTTTTGTGACGCGAGCGCGGTGTGTGTCAAC	1500
Db	1438	GAAAGCTCTGGGTGGAATACTCTCGCCGCTCTCTCTGACGCGTCTGGCGGTGTCTTTCATAC	1497
OY	1501	GCGCGGACCGAACCGGTATTTCCGATGCCCTTCAGCGAGGAAAAACGGCGGAGACTTGTGCGAG	1560
Db	1498	GCGGCGACGCAATCGCTACTTCCGTGCTCTCAGCGAGGAAATGGCGAGAACCCGTGTGGCAG	1557
OY	1561	GCCCGCTTGGGACGCGTGGCGAACGGGCGAGCGGATCACTTACGAGTTTGAATGGCGTGCAC	1620
Db	1558	ACCGGTCGGGACATGTGTGCCAGCGGTCAACCATCAAGCTACGACATGGAGCGGCTGCAG	1617
OY	1621	TATATCGGCATCGGTGCGGGCGGTCTGACCTATATGACACGCAATTGAAACGCGCGCCTGGCC	1680
Db	1618	TATGTGTCATCGCAGCGCGGCGGTATATCTTACGCGCACTAACGTAACAGCAATATGCGC	1677
OY	1681	GAGGCAATCGATTCGACTCGGTGGGTGATATCGCATCTATGTCTTTGCACTGCGCGGCACTAA	1740
Db	1678	GCGACATCGATTCGACTTCGATCGGCAACGCGGTCTACGCTTTCGGCCCTTTCGCGCAATTA	1737

RESULT	8
AD134122	ID
AD134122 standard; DNA; 2519 BP.	
AD134122;	AC
15-APR-2004 (first entry)	DT
Keroglomonicigenum sp. sorbicol dehydrogenase (SDH) 1 gene.	DE
Sorbicol dehydrogenase; SDH; cytochrome c; molecular biology/	KM
bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gu	KM
2KUG; gene; ds.	XX
Keroglomonicigenum.	OS
Location/Qualifiers	XX
741..744	FT
/*tag= a	FT
/standard_name= "Shine-Dalgarno"	FT
750..2486	FT
/*tag= c	FT
/product= "SDH protein"	FT
750..818	FT
/*tag= b	FT
819..2483	FT
/*tag= d	FT
/product= "Mature SDH protein"	FT

XX	US2003228672-A1.
PN	
XX	
PD	11-DEC-2003.
XX	
PF	06-JUN-2002; 2002US-00162713.
XX	
PR	06-JUN-2002; 2002US-00162713.
XX	
PA	(ARCH) ARCHER-DANIELS MIDLAND CO.
PI	
P1	Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF;
XX	Yum D;
DR	WPI; 2004-052025/05.
PT	P-PSDB; ADI34118.
XX	
PS	New proteins of sorbitol dehydrogenases and cytochrome c of the strains
CC	Ketogulonigenium spp., useful in molecular biology, bacteriology and
CC	industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
CC	gulonic acid.
CC	Claim 37; SEQ ID NO 9; 68pp; English.
XX	
XX	The invention relates to the identification and isolation of nucleic acid
CC	sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
CC	of the strains, Ketogulonigenium sp. The proteins and nucleic acid
CC	molecules are useful in the fields of molecular biology, bacteriology and
CC	industrial fermentation specifically for producing L-sorbose and 2-keto-L-
CC	gulonic acid (KLG). The present sequence is Ketogulonigenium sp. SDH
CC	gene.
XX	
SQ	Sequence 2519 BP; 542 A; 797 C; 703 G; 477 T; 0 U; 0 Other:
Query Match	61.6%; Score 1071.2; DB 12; Length 2519;
Best Local Similarity	76.4%; Pred. No. 9.3e-259;
Matches 1329; Conservative	0; Mismatches 408; Indels 3; Gaps 1;
OY	1 ATGAACCCCAAGCGTCGCTTTCGCACAGCGCGCGGTGCTATTGCTTACCGGCCGCC 60
DB	750 ATGAATTCGAATTTGGTTGCTTCGCGAAGGCTTGCTGCCGTGCATTTTGGCTGCC 809
OY	61 GCATTTCGCGAGGTAAACCCGATTAACCATGAACTGCTGCGAACCCGCCGCTGTGAA 120
DB	810 GCATTTCGCGATGTAAGCGCCCGCTCACAAGAAGTGTCTAATAAACCCGCCCGCGGAA 869
OY	121 TGATTTACTACGCGCGCAAACCAAGAAACTATATGCACTTCGCCCTTCGACCCAGTACT 180
DB	870 TGATTCAGCTATGCGCGCAACCAAGAAAACCTACCGCACTTCGCGCTGAACCAATCAC 929
OY	181 GCCGACAACGTTGGTTCAGTTGCAACTGATGCGGCCCGCGGAGATGAGCGCGGCGCGTA 240
DB	930 CCCGACAACGTCGCGCGACGCTGCACTGTCTGCGGCCCGCGGAGATGAAACCCCGGCTGTG 989
OY	241 CAGGTCAACGCGATGATTCATGATGCGGTATGTATCTGCGAAAACCCCGGTATGTATC 300
DB	990 CAGGTGACCCCGCTGATGTCACAGACGGGTATGTATCTGCGAACCCGCGCAATCATTT 1049
OY	301 CAGGCGCTGATATGCGCAACAGCGAGCTGATCTGGGAACAACCGCGCGCAACTGCCGCC 360
DB	1050 CAAGCGATGACGCGCAAAACCGGTGACTTGTCTGGGAACAACCGCGCGCAACTGCCGAG 1109
OY	361 GTCCCAACGCTAAACGCCCAAGCGCAACCGCGCGCGTGTGCCCTTTACGCGACGAGC 420
DB	1110 ACCTTCAGACGCTCAGCTGCTGGGGGATTCGAAGGCGGCACTCGGCTTTATGGACCAAT 1169
OY	421 CTCTATTTCACTATGCGAACAACATCTGATGCGCGTGTGATATGAGAGAGGGGCAAGTC 480
DB	1170 GTCTACTTCGCTCTGTGGGAACAACAATGATGCGCTGTGATGCTGCAAGGGGCAAGTC 1229
OY	481 GTATTTCATGTCGAACGCTGATCGGCGGAAGACGCTTGAACCAATTAACACACCGGGCG 540
DB	1230 GTCTTCGACGTGACCCCGCGCAAGGAGAGAGGGGTCTCG--AACTGTGTCGGGCCC 1286


```

QY 541 ATTGTCGCAATGCGCTCATGCTCGGGGTTTCCACTCTGCCAATATTGCCCTATGATGC 600
DB 1287 ATTTGGCCCAACGGCGGATCGTGTGCGGTTTCGACCTGCCAATATCGCCCTTCGGCTGT 1346
QY 601 TTATATCTCGGGGACGATTCGCGACGGGATGAGAGCTGTGGCGCAACCTTTATCCG 660
DB 1347 TTTGTGTGCGGCCATGATGAGACGCGGCGAAGAACTGTGGCGCAACTTATCTTCACTCCG 1406
QY 661 CAGCGCGGCGAAGAGGCTGACGAGACTTGGGGCAATGATTTGAGGCGCGCTTGATGACC 720
DB 1407 CAGGAGGTGAAGAGGCTGACGAAACCTGGGGCAATGATTAAGAAACCCGCTGATGACC 1466
QY 721 GGGGCTGGGGTCAATCACTATGATCCGCTGACGAACTGTGTTTATATGCTGACG 780
DB 1467 GGGGCTGGGGCCAGATCACTATGATCCCACTATGATTTGGTATTTTACGATGCTG 1526
QY 781 GGGGTTGGCCCAAGCTCGAAGCCAGCGCGGACCGCGGCGACGCTGTATGCGACC 840
DB 1527 GCGGTAGGCCCGGCAATCCGAGTTTCAGCGCGGACCCCGGCGGCAOGCTTTACGCGACC 1586
QY 841 AAGACCCGCTTTGCGGTGCTCCCGACACGGGCGAATTTCTGGGCTCAACGAGACCTTG 900
DB 1587 AAGACCCGCTTTGCGGTGCTCCCGACACGGGCGAAGTGTCTGGGCTCAACGAACTTG 1646
QY 901 CCGCGCGCAACTGGGACCAAGATGCACTTTCGAGATGATGGTCCGCAACGTCGATG 960
DB 1647 CCGCGCGCAACTGGGACCAAGATGCACTTTCGAGATGATGGTCCGCAACGTCGATG 1706
QY 961 CAACCTCGGCGGAGATGAGGCTCTGCGCGCATCAACCCATGCGCGACGCGCGAG 1020
DB 1707 CAGCGCGCTGCGCAATGAGCGCGGCGCAAGCCATCAACCCATGCGCGACGCGCGAG 1766
QY 1021 CGCGGTGTGCTGACGGGTGCGCTTGGCAAGCCGCGACGATGCTGTTATGCGGCGC 1080
DB 1767 CGTGTGCTTGTACCGCGCTTCCGTGCAAAACCGGTACATGTGCGAGTTTCAGCTGAA 1826
QY 1081 TCGGGCGAATTCGTGTGGGCGCGTGTATACCACTAGACCAATATGATGCGCTGATGAC 1140
DB 1827 ACGGGCGAATTCGTGTGGGCGCGCGACACCACTAGACCAATATGATGATGATGAC 1886
QY 1141 GAGACGCGCTTGTGACGCGTGAACGAGAGATGCGGTCTGTAAGAGCTGACGTTGAAT 1200
DB 1887 GAAACCGGTGTGTCAGCGTGAATGAATATCATCTTAAGATGTGACACCGACTAC 1946
QY 1201 GAGCTGTGCGGACCTTCTGGGTGCGCGCATGCTGTGTCAGCGCGACTGAAACCGGAC 1260
DB 1947 GCGATTTGCCGCAATTTCTTGGGTGACGCGACTGGGCGGTGCGCATCTTGAACCCGAT 2006
QY 1261 ACCGGCATTTACTTCTTGGCGGTGAACATGCTCTGCTAGATATTAAGCCGTTGATCAA 1320
DB 2007 AGGGGCAATCTACTTCAATTCCTCGAACAACGCTGTGCGAATTTGGCGGACGTCGATCAA 2066
QY 1321 GAGTTTAGCGGCTCGACGCTCTAATAACACAGCGCGACCGCAAAATCTGCGCGGCGCTTT 1380
DB 2067 GAGTTTAGCGGCTCGACGCTCTAATAACACAGCGCGCACTTACTGTTTGGCGCGGAAAA 2126
QY 1381 GAAATATGCGCGCATGACGCGATTAATATGACGCGCGGCGCACTTGTGTGTGCGG 1440
DB 2127 GAAATATGCGCGCATGACGCGATTAATATGACGCGCGGCGCAAAACCTGTGTGTGCGG 2186
QY 1441 GAGCGCTGCGGCGAATCTATGCGCGCTTGTGTGACGCGGCGGCTGTGTGTGATCAAC 1500
DB 2187 GAGCGCTGCGGCGAATCTATGCGCGCTTGTGTGACGCGGCGGCTGTGTGTGATCAAC 2246
QY 1501 GGGCGGACGACCGCTATTTCCGTGCGCTTACGCGAGAAACCGGCGAGACTTTGTGTGAG 1560
DB 2247 GGGCGGACGACCGCTATTTCCGTGCGCTTACGCGAGAAACCGGCGAGACTTTGTGTGAG 2306
QY 1561 GCGCGCTTTCGACGCGTGTGACGCGGCGGCGGCGATCACTGATGAGATTTGAGCGCGTGA 1620
DB 2307 ACCGCTGTGCGACTGTGTGCGACGCGGTCAAGCATAGCTGACGAGCGCGGTGAG 2366

```

```

QY 1621 TATATGCGCATTCGCTGCGGGGCTGTGACTTATGAGACGCAATGAACGCGCCCTGCC 1680
DB 2367 TATGTTGCGCATTCGCGGGGCGGTATATCTTACGCGCACTTACGAGCAATATGCGC 2426
QY 1681 GAGGCAATTCGATTCGACTCGGTGCGTATGCGATCTATGCTTTTGCAGCTGCCAGTAA 1740
DB 2427 GCGACCATCGATTCGACTTCGATCGGCAACGCGGTCTTACGCTTTCGCGCTTCCGATTA 2486

RESULT 9
AAV29052
ID AAV29052 standard; cDNA; 1740 BP.
XX
AC AAV29052;
XX
DT 21-AUG-1998 (first entry)
XX
DE Alcohol and/or aldehyde dehydrogenase A' nucleotide sequence.
XX
KW Alcohol/aldehyde dehydrogenase A' enzyme; recombinant organism; aldehyde;
KW ketone; carboxylic acid; L-sorbose; D-sorbitol; 2-keto-L-gulononic acid;
KW L-ascorbic; inhibition; ds.
XX
OS Gluconobacter oxydans.
XX
FH Key Location/Qualifiers
FT 1..1740
FT CDS
FT
FT /tag=a
FT /product="alcohol and/or aldehyde dehydrogenase A'"
XX
PN EP832974-A2.
XX
PD 01-APR-1998.
XX
PP 11-SEP-1997; 97EP-00115801.
XX
PR 19-SEP-1996; 96EP-00115001.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
PI Asakura A, Hoshino T, Ojima S, Shinjoh M, Tomiyama N.
XX
DR WPI: 1998-195228/18.
XX
DR P-PSDB; AAM37874.
XX
PT Recombinant Gluconobacter oxydans alcohol and/or aldehyde dehydrogenase
PT enzyme(s) - useful for converting substrate(s), e.g. L-sorbose or D-
PT sorbitol to 2-keto-L-gulononic acid.
XX
PS Disclosure: Page 29-30; 5pp; English.
XX
PS This is the nucleotide sequence encoding the Gluconobacter oxydans
CC Alcohol and/or aldehyde dehydrogenase A' enzyme. The enzymes or
CC recombinant organisms can be used to convert suitable substrates to
CC aldehydes, ketones or carboxylic acids, especially to convert L-sorbose
CC or D-sorbitol to 2-keto-L-gulononic acid, which can be converted to L-
CC ascorbic acid by standard procedures. The derivatives of ADH enzymes
CC have desired substrate specificity, higher affinity to a substrate, lower
CC affinity to an inhibitory compound, higher stability against temperature
CC and/or pH and higher catalytic speed
XX
SQ Sequence 1740 BP; 347 A; 576 C; 503 G; 314 T; 0 U; 0 Other;

Query Match 60.6%; Score 1053.6; DB 2; Length 1740;
Best Local Similarity 75.7%; Pred. No. 2.2e-254;
Matches 1318; Conservative 0; Mismatches 419; Indels 3; Gaps 1;

QY 1 ATGAACCCCAACAACGTCCTTGCACACGCGGCGGTATATGCTTACCGGCGCGGCC 60
DB 1 ATGAAGACGTGCTCTTTCCTGCTGAGCGCTTGCAGCTTGAAGCTATAGTCTCTTT 60
QY 61 GCATTCGCGAGGTAAACCCGATTAACGATGACTGCTGAGGAAACCGCGCTGTGAA 120

```

Db 61 GGCGTTGCTCAAGTGAACCCCGTACACGATGAAATGCTGGCGAACCCCGCGCTGGTGA 120
 QY 121 TGAATTAACAGCCCGGCAACCAAGAAACTATCCGCACTCCGCCCTGACCAAGATCACT 180
 Db 121 TGAATTAACAGCCCGGCAACCAAGAAACTATCCGCACTCCGCCCTGACCAAGATCACT 180
 QY 181 GCCGACAAAGTGTGAGTGAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 240
 Db 181 ACTGAGAAAGCTCGGCGCACTGCAACTGCTGTGGCGCGCGGCAATGAGCGCGGCAAAATC 240
 QY 241 CAGGTCAACCGCATGATCCATGATGCGGTGATGATGATGATGATGATGATGATGATGAT 300
 Db 241 CAGGTCAACCGCATGATCCATGATGCGGTGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 CAGGCGGTGATGCGGAAACAGGCACTGATCTGGGAAACACCGCGGCAACTGCGCGG 360
 Db 301 CAGGCGGTGATGCGGAAACAGGCACTGATCTGGGAAACACCGCGGCAACTGCGCGG 360
 QY 361 GTGCGCAACGCTAAACCGCCAAAGCGACCGCAAGCGCGCGTGTGCTTTACGCGCACAGC 420
 Db 361 GTGCGCAACGCTAAACCGCCAAAGCGACCGCAAGCGCGCGTGTGCTTTACGCGCACAGC 420
 QY 421 CTGATTTTACGCTCAAGGCAACACATGATGCGGTGATGATGATGATGATGATGATGATGAT 480
 Db 421 CTGATTTTACGCTCAAGGCAACACATGATGCGGTGATGATGATGATGATGATGATGATGAT 480
 QY 481 GTATTCGATGCGAAGCGTGAATCGGCGAAGACGCGCTTGAACGATTAACACACGCGGCG 540
 Db 481 GTATTCGATGCGAAGCGTGAATCGGCGAAGACGCGCTTGAACGATTAACACACGCGGCG 540
 QY 541 ATTGTGCGCAATGCGCTGATCTGCGGGGTTCACCTGCGCAATATTCGCGCTGATGATGATG 600
 Db 541 ATTGTGCGCAATGCGCTGATCTGCGGGGTTCACCTGCGCAATATTCGCGCTGATGATGATG 600
 QY 598 TTTATCTCGGCGACGATTCGCGAAGCGTGAATCGGCGAAGACGCGCTTGAACGATTAACAC 657
 Db 598 TTTATCTCGGCGACGATTCGCGAAGCGTGAATCGGCGAAGACGCGCTTGAACGATTAACAC 657
 QY 661 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 720
 Db 661 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 720
 QY 721 GGCGTGTGGGTGAGATCACTATGATCCGCTGACGAACTTGTGTTCTATGAGCTGACG 780
 Db 721 GGCGTGTGGGTGAGATCACTATGATCCGCTGACGAACTTGTGTTCTATGAGCTGACG 780
 QY 778 GGTGTGCGCGGCTTGGAAACCCAGCGCGCACACCGCGCGCACTGATGATGATGATGATG 837
 Db 778 GGTGTGCGCGGCTTGGAAACCCAGCGCGCACACCGCGCGCACTGATGATGATGATGATG 837
 QY 841 AAGACCGGCTTTCGCGTCCGCGCAACCGCGGCAATGCTGCGGTGATGATGATGATGATG 900
 Db 841 AAGACCGGCTTTCGCGTCCGCGCAACCGCGGCAATGCTGCGGTGATGATGATGATGATG 900
 QY 897 AAGACCGGCTTTCGCGTCCGCGCAACCGCGGCAATGCTGCGGTGATGATGATGATGATG 957
 Db 897 AAGACCGGCTTTCGCGTCCGCGCAACCGCGGCAATGCTGCGGTGATGATGATGATGATG 957
 QY 961 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 1020
 Db 961 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 1020
 QY 1017 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 1077
 Db 1017 CAGCGCGGCGAAGCGGTGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGATGATG 1077
 QY 1081 TCGGCGGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATG 1140
 Db 1081 TCGGCGGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATG 1140
 QY 1137 ACCGCGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1197
 Db 1137 ACCGCGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1197

QY 1201 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1260
 Db 1201 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1260
 QY 1257 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1317
 Db 1257 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1317
 QY 1317 ACCGCGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1377
 Db 1317 ACCGCGCAATTCCTGTGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1377
 QY 1377 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1437
 Db 1377 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1437
 QY 1437 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1497
 Db 1437 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1497
 QY 1497 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1557
 Db 1497 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1557
 QY 1557 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1617
 Db 1557 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1617
 QY 1617 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1677
 Db 1617 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1677
 QY 1677 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1737
 Db 1677 GAGCTTGCCGACCTTCTGAGTGGCGCGAGATGATGATGATGATGATGATGATGATGATGAT 1737
 RESULT 10
 ADI34116
 ID ADI34116 standard; DNA; 1740 BP.
 XX
 AC ADI34116;
 XX
 DT 15-APR-2004 (first entry)
 XX
 DE Ketogulonidigenium sp. sorbitol dehydrogenase (SDH) 2 DNA.
 XX
 KW Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
 XX bacteriology; industrial fermentation; L-sorbitose; 2-keto-L-gulonidic acid;
 XX 2KLG; gene; ds.
 CS Ketogulonidigenium.
 XX
 FH Key
 FT CDS 1.1740 Location/Qualifiers
 FT FT /*tag= b
 FT FT /product= "SDH protein"
 FT sig_peptide 1..69
 FT FT /*tag= a
 FT FT mac_peptide 70..1737
 FT FT /*tag= c
 FT FT /product= "Mature SDH protein"
 PN US200328672-A1.
 XX
 PD 11-DEC-2003.
 XX
 PF 06-JUN-2002; 2002US-00162713.
 XX
 PR 06-JUN-2002; 2002US-00162713.
 XX
 PA (ARCH) ARCHER-DANIELS MIDLAND CO.

XX Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF,
PI Yun D;
XX WPI; 2004-052025/05.
DR P-PSDB; ADI34120.
XX
PT New proteins of sorbitol dehydrogenases and cytochrome c of the strains
PT ketoglulonigenium spp.; useful in molecular biology, bacteriology and
PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
PT gulonic acid.
PS Claim 21; SEQ ID NO 3; 68bp; English.
XX
CC The invention relates to the identification and isolation of nucleic acid
CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
CC of the strains, Ketoglulonigenium sp. The proteins and nucleic acid
CC molecules are useful in the fields of molecular biology, bacteriology and
CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-
CC -gulonic acid (2KLG). The present sequence is Ketoglulonigenium sp. SDH
CC DNA.
XX
SQ Sequence 1740 BP; 369 A; 567 C; 494 G; 310 T; 0 U; 0 Other;
Query Match 60.5%; Score 1052; DB 12; Length 1740;
Best Local Similarity 75.7%; Pred. No. 5.ee-254;
Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;

DB 658 AAAGCGGGTGAAGAGGAGATGAACCTGGGGCAACGACTGAGAGCCCGCTGATGACC 717
QY 721 GGCGCTGGGGGTCAAGATCACTATGATCCGGTGAAGAACTTGTGTTCTATGCTGACAC 780
DB 718 GGCGCTGGGGGCAAAATACAGTACGACCCCGTCAACCACTGTATTTCTACGATCGTCG 777
QY 781 GGCGTGGGCGCCAGCGTCCGAAACCCAGCGGGCAACCGCGGGCGGACCTGTATGGCAC 840
DB 778 GCGGTGCGCGCGCTTTCGAAACCCAAAGCGGACCAACCGGCGGACCATGTACGGCACG 837
QY 841 AACACCCGCTTTCGAGTCCGATCCGACACGCGGCGAGATTTGTCTGCGCTCACACAGCCCTG 900
DB 838 AACACCCGCTTTCGCGTCCGTCGCGCCGACACCGCGGAAATCGTCTGGCGTCAACAACTCTG 897
QY 901 CCGCGCAAACTGGGACCAAGAAATGACGTTCCGATGATGTGTCCCAACGTGATGTG 960
DB 898 CCCCCTGCAACTGGGACCAAGAGTGCAGTTCCGAAATGATGTGTCCCAATGTGACGCTC 957
QY 961 CAACCTCGGCGAGATGAGGGTTCGCGGCAATCAACCCCAATGCGGGGAGCGGAG 1020
DB 958 CAGCTTGGCTGACATGAGAGGCTGAAGTCAATCAACCCCAACCGCCACTGGCGAG 1017
QY 1021 CGCGCTGTGTGACGCGGTGCGCTTGCAGACCGGACGATGTGCTGTTGATGCGGCC 1080
DB 1018 CGTGCCTGTGTGACGCGGTTCGCGCAAAACCGGTACATGTGGCAGTTGACGCTGAA 1077
QY 1081 TCGGCGCAATTCCTGTGGGCGCGGTGATACCAATCAACCAATATGATGCTCGATGCAC 1140
DB 1078 ACGGGCAATTCCTGTGGGCGCGGACCAACCAATCAATCAATGATGATGATGATGAC 1137
QY 1141 GAGACCGGCTTGTGACGCGGTGAAGAGATGCGGTGCTGAAGAGAGCTGAAATAT 1200
DB 1138 GAAACCGGCTTGTGTACGCGGTGAATGAAGATATCAATCAATCAATCAATGATGATGAC 1197
QY 1201 GACGTCTGCGGACCTTCTCGGGTGGCGGAGTGTGTGTGACGCGACCTGACACCGGAC 1260
DB 1198 CGCATTTGCGCGCAATTTGTGGGTGAGCGACATGCGCGGTGCGGATCTTGTAAACCGGAT 1257
QY 1261 ACCGCACTTACTTCTTCCGCTGAACAATGCTGCTACGATATATATGAGCGGTGATCAA 1320
DB 1258 AGCGGCACTACTTCACTTCCCTGAAACAAGCGCTGTGCGGATTTGGGCGAGTGATCA 1317
QY 1321 GAGTTAGCGGCTGACGCTGTATPAACAACAGCGGACCGGCAAACTGCGCGGCGCTTT 1380
DB 1318 GAGTTCAAGCAATGAGAGCTTACACAACAGCGGACCTTACCTGCTTGGCGGAAAAA 1377
QY 1381 GAAATATAGGCGCGCATGAGAGGATGATATCAAGCAACCGGCGGACCTTGTGGCGG 1440
DB 1378 GAAATATAGGCGCGCATGAGAGGATGATATCAAGCAACCGGCAAACTGCTGTGTGCTC 1437
QY 1441 GAGCGCCCTGCGGGAATCTACCTGCGCTTGTGTGACAGCGGCGGCTGTGTGTTCAAC 1500
DB 1438 GAAAGTGTGGGCTGGAATCTACCTGCGGCTTGTGACAGCGGCGGCTGTGTGTTCAAC 1497
QY 1501 GCGGGAACCAACGCTATTTCCGTGCTTCAAGCAAGAAACCGCGGACCTTGTGGCAG 1560
DB 1498 GCGGGAACCAACGCTATTTCCGTGCTTCAAGCAAGAAACCGCGGACCTTGTGGCAG 1557
QY 1561 GCGGCTTGTGAGAGTGTGAGAGGCGGCGATGACGATGACGATGACGATGACGATGAC 1620
DB 1558 ACCGCTGTGGGAGCTGTGCTTCCGCGCAAGCGCTGTGTGACGATGACGATGACGAT 1617
QY 1621 TATATGCGATGCGGTGCGGCGGTGTGACCTATGAGCAACGCAATGAAACGCGCGCTGACC 1680
DB 1618 TACATGCGATGCGGTGCGGCGGTGTGACCTATGAGCAACGCAATGAAACGCGCGCTGACC 1677
QY 1681 GAGGCAATGCAATGCACTGCGGTGCGGTATGATGATGATGATGATGATGATGATGATGAT 1740
DB 1678 GAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737

RESULT 11

```

AD134123
ID AD134123 standard; DNA; 3200 BP.
XX AC AD134123;
XX DT 15-APR-2004 (first entry)
XX DE Ketogulonigenium sp. gene encoding cytochrome c and SDH proteins.
XX KW Sorbitol dehydrogenase; SDH; cytochrome c; molecular biology;
XX KM bacteriology; industrial fermentation; L-sorbose; 2-keto-L-gulonic acid;
XX 2KLG; gene; ds.
XX OS Ketogulonigenium.
XX FH Key Location/Qualifiers
XX FT RBS /tag= a
XX FT CDS /standard_name= "Shine-Dalgarno"
XX FT /tag= c
XX FT /product= "Cytochrome c protein"
XX FT sig_peptide /tag= b
XX FT mat_peptide /tag= d
XX FT RBS /product= "Mature cytochrome c protein"
XX FT /tag= e
XX FT /standard_name= "Shine-Dalgarno"
XX FT CDS /tag= g
XX FT sig_peptide /product= "SDH protein"
XX FT mat_peptide /tag= f
XX FT /tag= h
XX FT /product= "Mature SDH protein"
XX US2003228672-A1.
XX 11-DEC-2003.
XX 06-JUN-2002; 2002US-00162713.
XX 06-JUN-2002; 2002US-00162713.
XX (ARCH) ARCHER-DANIELS MIDLAND CO.
XX PA Choi E, D'elia J, Kim H, Kim M, Lee JK, Pan J, Stoddard SF;
XX PI Yum D;
XX DR WPI; 2004-052025/05.
XX DR P-PSDB; AD134120.
XX PT New proteins of sorbitol dehydrogenases and cytochrome c of the strains
XX PT Ketogulonigenium spp., useful in molecular biology, bacteriology and
XX PT industrial fermentation, e.g. for producing L-sorbose and 2-keto-L-
XX PT gulonic acid.
XX PS Claim 45; SEQ ID NO 10; 68bp; English.
XX CC The invention relates to the identification and isolation of nucleic acid
XX CC sequences and proteins of sorbitol dehydrogenases (SDH) and cytochrome c
XX CC of the strains, Ketogulonigenium sp. The proteins and nucleic acid
XX CC molecules are useful in the fields of molecular biology, bacteriology and
XX CC industrial fermentation specifically for producing L-sorbose and 2-keto-L-
XX CC gulonic acid (2KLG). The present sequence is Ketogulonigenium sp. gene
XX CC encoding cytochrome c and SDH proteins.
SQ Sequence 3200 BP; 711 A; 1018 C; 880 G; 591 T; 0 U; 0 Other;
Query Match 60.5%; Score 1052; DB 12; Length 3200;

```

```

Best Local Similarity 75.7%; Pred. No. 6,6e-254;
Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;
QY 1 ATGAACCCCAACGCGCTTCGACACGAGCGCGCGTGTATGCTACCGCGCCGCC 60
DB 1241 ATGAAGACGAAGTCTTTCTGTTGTCAGAGCGCTTCTGCTGCAAGCTACGACCAATT 1300
QY 61 GCATTGCGGAGGTAACCCGATTAACGATGAATCTGTGCGAACCCTGCTGTGAA 120
DB 1301 GCGCTTCTGATGACCCCGTACCGACGACTGTGCAAAACCCGCGCGCGAA 1360
QY 121 TGGATTAACTACGCGCGCAACCAAGAAATACTATGCGACCTGCGCCCTGACCGACAT 180
DB 1361 TGGATCAGCTACGCGCGCAACCAAGAAATACTATGCGACCTGCGCCCTGACCGATCAG 1420
QY 181 GCCGACACGTTGGTCACTGTTGCACTGCTGCGCGCGCGGAGTGAAGCGGCGGTA 240
DB 1421 CCGAGAACGTCGCTGACGTCGCACTGCTGCGCGCGCGGATGAACCGCGCAAGTTC 1480
QY 241 CAGGTACCGCGATGATTCATGATGCGGTATGATTTGCGCAAAACCCGCTGATGATG 300
DB 1481 CAAGTCACTCGCGTGAATCAATGATGCGGTATGATTTGCGCAAAACCCGCGATGATC 1540
QY 301 CAGCGCTGATGTCGCAACAGGCGATCTGATGCGCAACCGCGCGCAACTGCGCGC 360
DB 1541 CAAGCATCGACGCTAAACCGGAGACTGATCTGCGCAACCGCGCGCGCGCGCAAC 1600
QY 361 GTGCCACGCTAAACGCGCAAGCGAGCGCAAGCGCGCGCTGCGCTTACGCGACAGC 420
DB 1601 GTGCCACGCTAAACGCGCAAGCGAGCGCGCAAGCGCGCGCTGCGCTTACGCGACAGC 1660
QY 421 CTCATTTCAAGCTATGAGGCAACCATGATGATGCGCGCTGATGAGAGCGCGCGAGT 480
DB 1661 GTTACTTCTGCTGCTGAGGCAACCATGATGATGCGCGCTGATGAGAGCGCGCGAGT 1720
QY 481 GTATTCATGTCGAACGTCGATGCGGCGGAGAGCGCTTGAACGATACACACGCGCGC 540
DB 1721 ACGTTCAGCTGCAACCGCGCGCAAGCGGAGAGAGATGCTTC--TAACCTGTCGCGCGC 1777
QY 541 ATTTCGCCAATGCGCTGATGTCGCGGCTTCAACCTGCGCAATTTGCGCTTATGATG 600
DB 1778 ATGTCGCTAACGCGGATGATGTCGCGGCTTCAACCTGCGCAATTTGCGCTTATGATG 1837
QY 601 TTATCTCGGCGCAAGTTCGCGCGAGCGGTCGAGAGCTGTCGCGCAACCTTATTCG 660
DB 1838 TTGCTTCGCGCGCAAGTTCGCGCGAGCGGTCGAGAGCTGTCGCGCAACCTTATTCG 1897
QY 661 CAGCGCGCGCAAGTTCGCGCGAGCGGTCGAGAGCTGTCGCGCAACCTTATTCG 720
DB 1898 AAACCGGTCGAAGAGCGGATGAACCTGCGCGCAAGCTGTCGCGCAACCTTATTCG 1957
QY 721 GCGCTGCGGCTGATCACTATGATTCGCGCGCAAGCTTGTGTCATGCTGCGAC 780
DB 1958 GCGCTGCGGCTGATCACTATGATTCGCGCGCAAGCTTGTGTCATGCTGCGAC 2017
QY 781 GCGCTGCGGCTGATCACTATGATTCGCGCGCAAGCTTGTGTCATGCTGCGAC 840
DB 2018 GCGCTGCGGCTGATCACTATGATTCGCGCGCAAGCTTGTGTCATGCTGCGAC 2077
QY 841 AAACCGGCTTGTGCGGCTGCGCGCAAGCTTGTGTCATGCTGCGAC 900
DB 2078 AAACCGGCTTGTGCGGCTGCGCGCAAGCTTGTGTCATGCTGCGAC 2137
QY 901 CCGCGCGCAACTGCGCAACGAATGACGCTTGTGATGATGCTGCGCAAGCTGATG 960
DB 2138 CCGCGCGCAACTGCGCAACGAATGACGCTTGTGATGATGCTGCGCAAGCTGATG 2197
QY 961 CAACCTTGTGCGGAGTGAAGCGGCTGCGCGCAACGAATGATGATGCTGCGCAAGCTG 1020
DB 2198 CAGCTTGTGCGGAGTGAAGCGGCTGCGCGCAACGAATGATGATGCTGCGCAAGCTG 2257
QY 1021 CGCGCTGTGCTGACGCGGTCGCTTGTGCAAGCGCGCAAGTGTGCTTGTGATGCGGC 1080

```

Db 2258 CGTGGCGTGTGACCGGCGGTTCCTGCGAANAACCGGTACATGTGCGAGTTGACGCTGAA 2317
 Qy 1081 TCGGGCGAATTCTCTGGGCGCGGTGATACCACTACGAAATGATGCGCTGATGAC 1140
 Db 2318 ACGGGCGAATTCCTGTGGGCGCGGACACCAACTACCAAAACATGATCACTTGATGAC 2377
 Qy 1141 GAGACCGGCTTGTGACGCTGAACGAGATGCGGTGCTGAAGAAGCTGACGTTGATAT 1200
 Db 2378 GAAACCGGCTGTGCTGACGCTGAATGAGATATCATCTTAAAGATCTGACACCGACTAC 2437
 Qy 1201 GAGCTCGCCGACCTTCTGGGTGGGCGGACGCTGTCTGACGCGACTGAACCCGGAC 1260
 Db 2438 CGCATTTGGCCGATCTTGTGGGTGAGCGGACTGCGCGTGGCATCTTGAACCCCGAT 2497
 Qy 1261 ACCGCGATTACTTCTTGGCGCTGAACCATGCGCTGCTACGATATATATGCGCTTGAACA 1320
 Db 2498 AGCGGCACTACTTCAATCCCTGGAACAACGCGCTGTGCGGATTTGGCGGCGAGTCATCA 2557
 Qy 1321 GAGTTAGCGCGCTCGACGCTTATACACACGCGCGACCCGAAAACCTCGCGCGGCTTT 1380
 Db 2558 GAGTTACGCGCAATGACGCTTACACACACGCGCGACTTACCTGCTTGGCGGAAAA 2617
 Qy 1381 GAAATATGGGCGCGCATGACGCGATGATATACACACGCGCGCGACTTGTGGCGCG 1440
 Db 2618 GAAATATGGGCGCGCATGACGCGATGATATACACACGCGCGCGACTTGTGGCGCG 2677
 Qy 1441 GAGCGCCCTGGCGGCACTACTGCGCCGTTTGTGACGCGCGCGGTGTGTTCAAC 1500
 Db 2678 GAGCGCTGTGCGCTGAACTACTGCGCGCTTCTGACGCGCGCTGCGCGGTGTGTTCAAC 2737
 Qy 1501 GCGCGGACCGACGCTATTTCCGTCGCTGACGCGGAAACCGCGGAGATTTGTGGCGAG 1560
 Db 2738 GCGCGGACCGACTGCTTCTTCCGCTGACGCGGAAACCGCGGAGATTTGTGGCGAG 2797
 Qy 1561 GCGCGTGTGCGACGCTGCGGACGCGGCGACGCGATCAGTACGATGTGACGCGCTGCA 1620
 Db 2798 ACCGCTGTGCGGACTGTGCTTGGGCGCAAGCGGTGCGTGAACGCGCGCTGCA 2857
 Qy 1621 TATATGCCCATGCTGTGCGGCGCTGTGACCTATGACGACGCAATTGAACGCGCGCTGCGC 1680
 Db 2858 TATATGCCCATGCTGTGCGGCGGACCACTACGCGCGGTGCGCAACGCTGCGCTGCGC 2917
 Qy 1681 GAGCGCATGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
 Db 2918 GAGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2977

RESULT 12

ADE94132
 ID ADE94132 standard; DNA; 1827 BP.

AC ADE94132;

DT 12-FEB-2004 (first entry)

DE Alcohol/aldehyde dehydrogenase coding sequence, SEQ ID 2.

KM Enzyme; alcohol/aldehyde dehydrogenase; pyridoxal quinoline quinone;
 L-sorbose; L-sorbose; 2-keto-L-gulononic acid; gene; ds.

OS Pseudoglucobacter saccharoketogenes; IF014464.

FH Key Location/Qualifiers

FT CDS 1..1827
 FT /tag= a
 FT /product= "Alcohol/aldehyde dehydrogenase"

JF2003159079-A.

PD 03-JUN-2003.

PF 29-NOV-2001; 2001JP-00364508.

XX

PR 29-NOV-2001; 2001JP-00364508.
 XX
 PA (FUT) FUISAMA PHARM CO LTD.
 XX
 DR WPI, 2003-818681/77.
 DR F-PSDB; ADE94131.
 XX
 PT Novel alcohol/aldehyde dehydrogenase protein catalyzing oxidation of
 PT hydroxymethyl group of compound to aldehyde group and aldehyde group of
 PT compound to carboxyl group, useful for manufacturing 2-keto-L-gulononic
 acid.
 PS Claim 6; SEQ ID NO 2; 48bp; Japanese.
 XX
 CC The present invention relates to an alcohol/aldehyde dehydrogenase (I;
 CC ADE94131), which catalyzes oxidation of hydroxymethyl group of a compound
 CC to an aldehyde group and the aldehyde group of a compound to a carboxyl
 CC group. (I) does not contain hemiterium and rare earth elements. (I) has
 CC an optimum pH of 4.5-5.5, isoelectric point of 4.1 +/- 0.3, pyridoxal
 CC quinoline quinone as a prosthetic group and K_m value of 40 mM for
 CC sorbose. (I) is useful for manufacturing a compound having a carboxyl
 CC group by contacting the compound having a hydroxymethyl group or an
 CC aldehyde group with (I) where the compound having a hydroxymethyl group
 CC or an aldehyde group is L-sorbose or L-sorbose and the compound having
 CC a carboxyl group is 2-keto-L-gulononic acid.
 XX
 SQ Sequence 1827 BP; 342 A; 612 C; 551 G; 322 T; 0 U; 0 Other;

Query Match 36.0%; Score 625.8; DB 10; Length 1827;
 Best Local Similarity 61.8%; Pred. No. 6.2e-147;
 Matches 1066; Conservative 0; Mismatches 647; Indels 12; Gaps 4;

Qy 24 CACACGCGCGCTGCTATTTGCTTACCGCGCCGCGCATTCGCGAGTAACCCGAT 83
 Db 99 CACCGCCCTGCGCGGACCGGACCAAGCGCGGCAAGCGCTATCGAAGACTTCAAGCCGT 158
 Qy 84 TACCGATGAATGCTGCGGGAACCGCGCGGTGTAATGATTAATCAACCGCGGCAACA 143
 Db 159 CAGGCTGACATGCTGCGGGAACCGCGGCAACCGCGGCAACCGCGGCAACCGGTA 218
 Qy 144 AGAAATATGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 203
 Db 219 CAGGCGTGGGATTAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 278
 Qy 204 ACTGCTGCGGCGCGCGGATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 263
 Db 279 GCTGCTGCGGCGCGCGGATGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 338
 Qy 264 TGGCGTGAATGATGCGCAACCGCGGATGATGATGATGATGATGATGATGATGATGATG 323
 Db 339 CGGCTGATCTTCTGCGGCAACCAACGATGATGATGATGATGATGATGATGATGATGATG 398
 Qy 324 GATCTGATCTGGAACAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
 Db 399 TTCCCTCATCTGGAATATGCTGCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 458
 Qy 381 AGGCGACCGCAAGCG 440
 Db 459 CGGCGCGCGCAAGCG 518
 Qy 441 CAACCATGATGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 500
 Db 519 CAATTCGTTGCG 578
 Qy 501 ATCGGCGCAACGCGCTGGAACAACCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
 Db 579 CAGGCGCTTAAAGAGCGCTGGAACCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 638
 Qy 561 CGTGGCGGTTCAACCTGCAATATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
 Db 639 GATGCGAGGCTCAACCTGCG 698
 Qy 621 CGGACGCGGTAGAGAGCTGCGCGCAACCACTTATCCCGCGCGCGCGCGCGCGCGCG 680

Db	639	TAAGTCGGGTGAAGAACTGTGGCGAACA	CTTCAATCCCCGCTCCGGCGAAGAGTGA	758
Qy	681	CGAGACTTGGGCAATG--ATTTCAGGCGCGCTGATGACCGGCGTCTGGGGTCA	GAT	737
Db	759	CGACACTCGGGGCGGCGCGCCCTACGAAAA	CCGTTGGATGACCGGTGCTTGGGCGCAAT	818
Qy	738	CACCTATGATCCCGTAGAAGAACTGTGTTCTA	TGCGTTCGACCGGCGTGGGCCGACGCGT	797
Db	819	CACCTATATCCGGAATTCGACTCGTGGTTTA	CTAGCGCTTCGACCGGCGCGGCCGCGCTTC	878
Qy	798	CGAAACCCAGCGCGGACGCGGGCGGACGCTG	ATGSCACCAACCCGCTTTGCGGT	857
Db	879	GGAAATCCAGCGCGGACCGAAGGCGGCACT	CTCGCAGAGCACAATACCCGCTTTCGCGT	938
Qy	858	GCGTCCCAACACGGGCGAGATTGCTTGGCGTCA	CCAGACCCCTCGCGCGGCAACTGGGA	917
Db	939	GAAAGCCAAAGCCGGTGAAGTGTCTGGAAGA	CCAGACCTGCGCCGCGCAACTGGGA	998
Qy	918	CCAAGATGACGTTCCGATGATGTCGCGCAAC	CTCATGTGCAACCTCGGCGCGAGAT	977
Db	999	CTCGGAATGCACTTTCGAATATATGTTGTTG	CTCGACGTCGCTGCAACCCGAGCCGCAAGC	1058
Qy	978	GGAGGCTGTGCGCGCATCAACCCCAATGCGG	CGACGGGCGA--GCGCGTGTGCTAC	1034
Db	1059	CGATGCAATGATGTCCTGTGCTGTCGTCGCA	AGTGGCGGACGAAACCCCAAGTGTGAC	1118
Qy	1035	GGGTCGCGCTTCGACAGCCGACGATGCGTGT	GATTGTCGGGCGCTCGGCGGAATTCCT	1094
Db	1119	CGGCGTGCCTGTCAGAACCGGCGTGCCTTG	GCAGTTTCGATGCAAGACCGGCGACTACTT	1178
Qy	1095	GTGGCGGCGTGAATACCACTACACCAATAT	GAATGCTCGATGACGAGACCGGCTTGT	1154
Db	1179	CTGTGTCAAAGGACACCGTGCAGACAGAA	CTCGATCGCTTCGATCGATGACACGGGCTGT	1238
Qy	1155	GACGCTGAACAGAGATGCGGTCGTGAAGA	CGTGGACGTTGAATGACGTTGCGCGGAC	1214
Db	1239	TACGCTCAATGAGCATGATCTCAAGGAG	CGGGCAAGACTTACATTACTGCGCGAC	1298
Qy	1215	CTTCTGAGGTGGGCGCGACTGTCGTCAGCG	CACTGAACCGGACACCGGCAATTA	1274
Db	1299	CTTCTCGCGGCTCGTGACTGGCGGTCGCG	CGGCTACGTCGGAAGTGAACCTTACGT	1358
Qy	1275	CTTGCCTGTAACAATGCTGCTACGATATTA	TGGCGTTGATCAAGTTTAGCGCGCT	1334
Db	1359	GATCCCGCTCAGCAACGCTGCTGACAGTGA	TGGCCGTTAGACCGAAGCCACTCGGAC	1418
Qy	1335	CGACGCTATTAACACAGCGGACCGGCAAA	ATTCGCGCGGCGCTTGGAAATATGCGCG	1394
Db	1419	TGACGCTTACACACCGACGACACGCTGT	GCTGCTCGGCGCAAGACCAATATGGTGC	1478
Qy	1395	CATCGACGGATTGATATACGACCGGCGCA	CTTGTGTCGGCGGACCGCTTCGGGC	1454
Db	1479	CGTGAATGCCATCGATCTCGCACCGGCGA	GCAGCAAGTGTCTACGAAACCGCTGCGGAC	1538
Qy	1455	GAACTATCTGCCCCGTTTTGTGCA	CGGACGCTGTGATGCTCAACCGGCGGACCGA	1514
Db	1539	TCTTACGACCGCGTCTGACCAACGCGGCG	CACTCGATTCGTCGCGGAGTATGATCG	1598
Qy	1515	CTATTTCGTCGCTCAGCGACGAAACCGG	CGCAGACTTGTGCGAGGCGCGCTTTCGAC	1574
Db	1599	TGACTTTCGCGCTGCGGACCGGAGCTCGG	GCAAGAGATCTGTGCCAACCGGCTC	1658
Qy	1575	GGTGCGACGGGCGAGGCGATACGCTACG	ATTGGAACGCGCTGCAATATATGCGCATCG	1634
Db	1659	TGCGGTTTCGCGCTACACACACGCTAC	TCAATGACATGACCGCGCATGTCGCGGTT	1718
Qy	1635	TGCGGCGCGTTCGACTATGAGACCGAAT	TGAACGCGCGCTGGCGGACGCAATGATTC	1694
Db	1719	CTCGGCGGCTGCTGCTGGTGGCCGACTT	---CGGCCGACCAACCGGAGTGACTC	1778
Qy	1695	GACTTCGCTCGGTAATGCAATATATGCTT	TTCGACTGCGCGACGTA	1739

Db 1776 GGGCTTCGGGCGGCAACGGCATCTACGTCTTCGCTTCCTCCGAGAA 1820

RESULT 13

ID	AAQ13580	standard; DNA; 2214 BP.
XX	AAQ13580;	
AC		
XX	27-ANG-2003 (revised)	
XX	25-VAR-2003 (revised)	
DT	09-DRC-1991 (first entry)	
XX		
DE	A.alcoetigenes membrane-bound ADH 72kd sub-unit.	
XX		
KW	alcohol dehydrogenase complex; carboxylic acid production; ss.	
XX		
OS	Acetobacter sp.	
PN	EP448969-A.	
XX		
PD	02-OCT-1991.	
XX		
EF	26-FEB-1991; 91EP-00102793.	
XX		
FR	26-FEB-1990; 90UP-00042391.	
XX		
FR	26-MAR-1990; 90UP-00073440.	
XX		
PA	(NAKA-) NAKANO VINEGAR CO LTD.	
PI	Tamaki T, Takemura H, Tayama K, Fukaya M, Okumura H, Kawamura Y;	
XX		
DR	WPI; 1991-289462/40.	
XX		
DR	P-PSDB; AAR13993.	
XX		
PT	Gene for membrane-bound alcohol dehydrogenase complex - obtd. from	
PT	Acetobacter alcoetigenes, used for prodn. of enzyme for converting	
PT	alcohol to acid.	
XX		
PS	Claim 3; Page 20-22 and Fig 3; 36pp; English.	
XX		
CC	Total DNA was prepared from A.alcoetigenes NH-24, digested with PstI and	
CC	SmaI and ligated to pSt-SmaI cleaved pUC18. The ligation mixture was	
CC	used to transform E.coli JM109. Probes were designed based on the N-	
CC	terminal amino acid sequence of the ADH complex isolated from A.	
CC	alcoetigenes (see AAQ13582-013584). The gene encoding the 72kd ADH	
CC	subunit was isolated and sequenced. It forms a membrane-bound ADH complex	
CC	with a 44kd sub-unit (cytochrome c) whose coding sequence is located	
CC	immediately downstream of the gene encoding the 72kd sub-unit. See also	
CC	AAQ13581. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 27-	
CC	AUG-2003 to correct OS field.)	
XX		
XX	Sequence 2214 BP; 451 A; 683 C; 662 G; 418 T; 0 U; 0 Other;	
SQ		
XX		
XX	Query Match 5.1%; Score 88.9; DB 2; Length 2214;	
XX	Best Local Similarity 50.8%; Pred.No. 4.9e-12;	
XX	Matches 266; Conservative 0; Mismatches 252; Indels 6; Gaps 2;	
QY	52 GGGCGCCCGCGCATTTGGCCGCAAGTAAACCCGATTACCGATGAACGTGGCGAAACCGGCC 111	
Db	97 GCATTCGGAGATGAGCGGCGAGGGCGCCACGGGGGAGACGATATCCATGCGATTATC 156	
QY	112 GCTGTGTAATGATTAACTACAGGCGCGCAACCAAGAAACTATCGCACTCGCCCTGACC 171	
Db	157 CCGGATATCGATATGACTATATGGCCGACACGTAATTCAGCAGGGCTAAAGCCCGGTGAT 216	
QY	172 CAGATCACTGCGCAACAAGTTGTCAGTTGCAATTTGGCTTGGGCCCGCGGAGATGAGGGG 231	
Db	217 CAGATCAACCGTTTCCAAATGTGCGTAACTGAAAGCTGGCCGTGATCTGACCTTGATACC 276	
QY	232 GGGGGCGGTACAGAGT---CAGCGCGATGATCATATAGGCTGATGTATCTGGCAACCCC 288	
Db	277 AACCGTGGCCAGGAAGCAGCCGCTGTATTATATGGCTCATATGACGACCAACCAAC 336	

CC subunit (subunit 1) of Gluconobacter oxydans sorbitol dehydrogenase
XX Sequence 2265 BP; 495 A; 670 C; 656 G; 444 T; 0 U; 0 Other;
SQ

Query Match 4.9%; Score 85.6; DB 2; Length 2265;
Best Local Similarity 50.3%; Pred. No. 3.1e-11;
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```
QY 100 GCGAACCCCGCGCTGCTGTAATGATTAATCTACGGCCGCAACCAAGAAAATATGCCCAC 159
    |||
DB 130 GCGACACAGCATCCGGTACTGATGAGTATGCGGACCTATTCCAGCAGGCTAC 189
QY 160 TCGCCCTGACCCAGATCACTGCCGACAGCTTGCTCACTGCTCACTGCTGCGCCGC 219
    |||
DB 190 AGCCCTGGATAGATCACTCAAGGACATGCGAGCACTGAAGTGGCATGSCACTAC 249
QY 220 GGGATGAGCGGGGGCCGCTACAGTC--ACGCCGATATCCATGATGCGTATAT 276
    |||
DB 250 GATCTGATACCAACCGTGTACAGGAAGGTACGCCGCTGATGCTGATGCGCTCATGTAC 309
QY 277 CTGGCAAAACCCGCGTATGATCCAGGCGTGGATCGCAACAGGCGATCTGATCTGG 336
    |||
DB 310 GCCACCAAACTGAGCAAGATGAAGCTCTGATGACGCTACGGGCAAGCTGCTGG 369
QY 337 GAACACCGCGCCCACTGCCCGCGTCCGCAAGCTAAACGCCCAAGCGA---CCGCAAG 393
    |||
DB 370 TCTTACGATCAAAAGGTTCCAGGCAACATGCGCACCGGCGTGGCTGATACGATCAAC 429
QY 394 CGCGCGCTGCGCTTTACGACGACGACCTCTATTTCAGCTCATGGGACAAACATCTGATC 453
    |||
DB 430 CGTGTGCAAGCTACTGAAAGGCAAAAGCTATTTCGACCTTCGACGCTGCTGATT 489
QY 454 GCGCTGATATGAGACGGGCGCAGTCTATTGATGTC-----GAAAGT 498
    |||
DB 490 GCCCTGATGCCAAGACCGGCAAGCTGTCTGAGAGCTCTATACGCTTCCAAAGGAAGG 549
QY 499 GGATCGGCGCAAGACGCTTGACCAATACACACGGGGCGATTGTCCCAATGGCGTC 558
    |||
DB 550 CAGCTGGTCAACAGCGCTCTACACGTTGACGCTGCCGTTACCCGTTACGCAAGGCAAG 609
QY 559 ATGCTGCGGGGTTCACTGCAATATTCGCCCTATGATGCTTTATCTCGGGGCAAGAT 618
    |||
DB 610 GTATCTCTGCGCAACGGCGGTGCAAGTTGCGGCGCCCGCTTGTGAGCGGTATGAC 669
QY 619 TCCGCGACGGGTAGAGAGCTGTGGCGAACCACTTATCCCGCAGCGGCGAAGGCT 678
    |||
DB 670 GCTGAACGGGAAAGATGAGCTGGCGCTTCTTACCGCTTCGAACCTTGACAAAGCGG 729
QY 679 GACG 682
    |||
DB 730 GACG 733
```

Search completed: November 22, 2004, 14:45:01
Job time : 949 secs


```

Db      301 CAGGCGCTGATGCGCAACAGGCGATCTGATCTGGGAAACACCGCCGCCAATGCGCCGCC 360
Qy      361 GTGCGACGCTAAACGCCCAAGCGACCGCAAGCGCGCGTGCCTTTTACGCGACGAGC 420
Db      361 GTGCGACGCTAAACGCCCAAGCGACCGCAAGCGCGCGTGCCTTTTACGCGACGAGC 420
Qy      421 CTCTATTTAGCTCATGCGCAACCATCTGATGCGCTGGAATATGGAACGCGGCCAGGTC 480
Db      421 CTCTATTTAGCTCATGCGCAACCATCTGATGCGCTGGAATATGGAACGCGGCCAGGTC 480
Qy      481 GATTTGATGATGCAAGTGGATGCGGGAAGACGCGTTGACGATTAACCAACGCGGCGC 540
Db      481 GATTTGATGATGCAAGTGGATGCGGGAAGACGCGTTGACGATTAACCAACGCGGCGC 540
Qy      541 ATTGTCGCAATGCGCTCATCTGCGCGGTTCCACTGCAATTTGCGCTTATGGAATGC 600
Db      541 ATTGTCGCAATGCGCTCATCTGCGCGGTTCCACTGCAATTTGCGCTTATGGAATGC 600
Qy      601 TTTATCTCGGGGACGATTCGCGGAGGAGGAGCTGCGGCAACCACTTTATCCG 660
Db      601 TTTATCTCGGGGACGATTCGCGGAGGAGGAGCTGCGGCAACCACTTTATCCG 660
Qy      661 CAGCGCGGCAAGAGGAGTGAAGACCTTGGGCAATGTTTGAAGCGCGCTGATGAC 720
Db      661 CAGCGCGGCAAGAGGAGTGAAGACCTTGGGCAATGTTTGAAGCGCGCTGATGAC 720
Qy      721 GCGGCTGCGGAGTCAATCACTTATGATCCCGTGAACCACTTGTGTTCTATGCGTGAAC 780
Db      721 GCGGCTGCGGAGTCAATCACTTATGATCCCGTGAACCACTTGTGTTCTATGCGTGAAC 780
Qy      781 GCGGTCGCGGACGCTCCGAAACCGACGCGGACCGCGGCGGACGCTGATGAGCAC 840
Db      781 GCGGTCGCGGACGCTCCGAAACCGACGCGGACCGCGGCGGACGCTGATGAGCAC 840
Qy      841 AACAACCGCTTTCGCGTCCGTCGCAACGCGGAGATTTGTCGCGTCAACGACCTG 900
Db      841 AACAACCGCTTTCGCGTCCGTCGCAACGCGGAGATTTGTCGCGTCAACGACCTG 900
Qy      901 CCGCGGCACTAGTGGACCAAGATGACGCTTGAATGATGATGATGATGATGATGATG 960
Db      901 CCGCGGCACTAGTGGACCAAGATGACGCTTGAATGATGATGATGATGATGATGATG 960
Qy      961 CAAACCTCGGCGGAGTGAAGGAGTTCGCGGCAATCAACCCCAATGCGCGGACG 1020
Db      961 CAAACCTCGGCGGAGTGAAGGAGTTCGCGGCAATCAACCCCAATGCGCGGACG 1020
Qy      1021 CGCGGTCGTCGAGCGGTCGCGCTTGAAGACGCGGACGATGATGATGATGATGATG 1080
Db      1021 CGCGGTCGTCGAGCGGTCGCGCTTGAAGACGCGGACGATGATGATGATGATGATG 1080
Qy      1081 TCGGCGCAATTCCTGTCGCGGAGTGAATCCCACTACCAATATGATGATGATGATG 1140
Db      1081 TCGGCGCAATTCCTGTCGCGGAGTGAATCCCACTACCAATATGATGATGATGATG 1140
Qy      1141 GAGACCGGCTTTCGTCGAGGAGTGAAGAGATGCGGTCGTAAGAGTGAAGCTGTAAT 1200
Db      1141 GAGACCGGCTTTCGTCGAGGAGTGAAGAGATGCGGTCGTAAGAGTGAAGCTGTAAT 1200
Qy      1201 GACGTCGCGGACCTTCTGCGGTCGAGGAGTGAAGCTGTCGACGCGCTGTAACCG 1260
Db      1201 GACGTCGCGGACCTTCTGCGGTCGAGGAGTGAAGCTGTCGACGCGCTGTAACCG 1260
Qy      1261 ACCGCGCATTTACTTTCGCGCTGAAACATGCTGCTGCTGATATATATGATGATCA 1320
Db      1261 ACCGCGCATTTACTTTCGCGCTGAAACATGCTGCTGCTGATATATATGATGATCA 1320
Qy      1321 GAGTTTACGCGCTGAGCTCTATTAACACAGCGGCAACGCAAACTGCGCGCGGCTTT 1380
Db      1321 GAGTTTACGCGCTGAGCTCTATTAACACAGCGGCAACGCAAACTGCGCGCGGCTTT 1380
Qy      1381 GAAAAATATGCGCGCATGACGCGATGATATGACGACCGCGGCGCACTTGTGTGCGG 1440
Db      1381 GAAAAATATGCGCGCATGACGCGATGATATGACGACCGCGGCGCACTTGTGTGCGG 1440

```

```

Qy      1441 GAGCGCTTCGCGGAACTACTTCGCGCTTTTGTGACGCGGCGGTGTGTTCAAC 1500
Db      1441 GAGCGCTTCGCGGAACTACTTCGCGCTTTTGTGACGCGGCGGTGTGTTCAAC 1500
Qy      1501 GCGCGGACCGACGCTATTTCCGTCGCTGACGCAAGAAACCGCGGAGACTTTGTGCGAG 1560
Db      1501 GCGCGGACCGACGCTATTTCCGTCGCTGACGCAAGAAACCGCGGAGACTTTGTGCGAG 1560
Qy      1561 GCCCGTTTCGACGCTGCGGACGCGGCGGATCACTGACGATTTGAACGCGCGCTGCGA 1620
Db      1561 GCCCGTTTCGACGCTGCGGACGCGGCGGATCACTGACGATTTGAACGCGCGCTGCGA 1620
Qy      1621 TATATGCGCATGCGTTCGCGGCGGCTGACCTATGACGCAATTTGAACGCGCGCTGCGC 1680
Db      1621 TATATGCGCATGCGTTCGCGGCGGCTGACCTATGACGCAATTTGAACGCGCGCTGCGC 1680
Qy      1681 GAGGCAATGATTCGACCTGCGTTCGATGATGATGATGATGATGATGATGATGATG 1740
Db      1681 GAGGCAATGATTCGACCTGCGTTCGATGATGATGATGATGATGATGATGATGATG 1740

```

```

RESULT 2
US-09-470-667-1
; Sequence 1, Application US/09470667
; Patent No. 6730503
; GENERAL INFORMATION:
; APPLICANT: Asakura, Akira
; APPLICANT: Hoshino, Tatsuo
; APPLICANT: Ojima, Setsuko
; APPLICANT: Shinjoh, Masako
; APPLICANT: Tomiyama, No. 6730503ibumi
; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases
; FILE REFERENCE: C38435/109700CON
; CURRENT APPLICATION NUMBER: US/09/470,667
; PRIOR APPLICATION NUMBER: US 08/934,506
; PRIORITY FILING DATE: 1997-09-19
; NUMBER OF SEQ IDS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Gluconobacter oxydans
US-09-470-667-1

```

```

Query Match      64.9%; Score 1128.6; DB 4; Length 1740;
Best Local Similarity 78.8%; Pred. No. 5.3e-288;
Matches 1373; Conservative 0; Mismatches 364; Indels 6; Gaps 2;

```

```

Qy      1 ATGAACCCCAACAGCTGCTTCGACCAAGCGCGCGTGTCTATTGCTTACCGGCGCGCC 60
Db      1 ATGAACCCGACTTTCGCTGCTTTGGGCGAGTGTGGCGCACTTGCACTTGTGCGCACCC 60
Qy      61 GCATTTCGCGAGTAACCCGATTAACGATGAACCTGTCGCAACCCCGCGTGTGA 120
Db      61 GCCTTTGCTCAAGGACCCCGTCAACCGATGAATGCTGCGAACCAGCGCGTGTGA 120
Qy      121 TGAATTACTACGCGCGCAACCAAGAACTATATGCGCACTGCGCCCTGACCGCATCACT 180
Db      121 TGAATGCTACGCTGCTGCAACCAAGAACTATGCTGCGCGCGCGCGCATGCACTG 180
Qy      181 GCGGCAACGTTGCTGAGTTCGCACTGCTGAGGCGCGGAGTGAAGGAGGCGGCGGTA 240
Db      181 ACTGGAACGTCGCGCAACCTGCACTGCTGAGGCGCGGAGTGAAGGAGGCGGCGGTA 240
Qy      241 CAGGTACGCGCGATGATCATGATGCGGATGATGATGATGATGATGATGATGATGATG 300
Db      241 CAGGTACGCGCGATGATCATGATGCGGATGATGATGATGATGATGATGATGATGATG 300
Qy      301 CAGGCGCTGATGCGGCAACAGCGGATGATGATGATGATGATGATGATGATGATGATG 360
Db      301 CAGGCGCTGATGCGGCAACAGCGGATGATGATGATGATGATGATGATGATGATGATG 360

```

QY 361 GTGCCACGCTAAACGCCCAAGGCGACCGCGAGCGGGCGTGGCCCTTTACCGGCAAGAC 420
 Db 361 ATGCCACGCTGAACAGCTTTGGCGAGCGGACCCCGGCGATGGCGCTGTACCGGACCAAC 420
 QY 421 CTCATTTCAGCTCATAGGACCAACCTGTATGCGCTGGATATGAGAAGCGGCGCAGTTC 480
 Db 421 GTTACTTGTGTTCGGGACCAACACCTGTGCGCTTGCACACCGCACTGCGCCCAAGTG 480
 QY 481 GTATTGATGTGAAGTGGATCGGGCGAAGACGCGTTCAGCAGTAAACCAACGCGGCGCG 540
 Db 481 AGCTTGACGTTCACCGCGCCCAAGGAGAAAGACATGCTTCG--NACTGTGGGCGCG 537
 QY 541 ATTGTGCGCAATGCGCTATCGTGGCGGTTCCACCTGCGCAATATTCGCTATGATGTC 600
 Db 538 ATGTGGCAAAAGCGCGTATCGTTGGCGGTTGACCTGCAATATCTGCGGTGGCTGCG 597
 QY 601 TTTATCTCGGGGAGACATTCGCGAGCGGTTGAGAGACTGTGGGCGCAACCACTTTATCCG 660
 Db 598 TTTGTCTCGGGCCACACTCGGCGACCGGTGAAGACTGTGGGCGCAACTTCACTTCACTCCG 657
 QY 661 CAGCCGCGGAGAAAGGAGTGAAGACCTTGGGCGAATGATTTGAGGCGCGCTGATGACC 720
 Db 658 CCGGCTGGCGAAGAGGTGATGAGCTTGGGCGCAAGATTAAGAACCCGCTTGGATGACC 717
 QY 721 GCGCTCTGGGCTCAATCACTATGATCCCGTGAAGCAACCTTGTGTCTATGAGCTTGACC 780
 Db 718 GGTGCTTGGGCGCAGATCACTATGACCCCGTCAACCACTTGTCACTACGCGCTGACC 777
 QY 781 GCGGTGGGCGCAGCTCCGAAACCCAGCGCGGCGCGCGGCGGCGCACTGTATGAGCAAC 840
 Db 778 GCTGTGGGCTCGCGCTCGAAACCCAGCGGCGACCCCGGCGGCGCGCTGTACGCGACG 837
 QY 841 AACACCCGCTTGGCGGTGCGTCCGACACGCGGCGAATGTCTGGGTGACCAAGACCTTG 900
 Db 838 AACACCCGTTTCCGCGGTGCTCTGACACGCGGCGAATGTCTGGGTGACCAAGACCTTG 897
 QY 901 CCGCGGAGCACTGGGACCAAGATGCAAGTTCGATGATGATGTCGCCCAAGCTGATGTC 960
 Db 898 CCGCGGAGCACTGGGACCAAGATGCAAGTTCGATGATGATGTCGCCCAAGCTGATGTC 957
 QY 961 CAACCTCGGCGAGATGAGAGGTCTGCGCGCCATCAACCCCAATGCGGCGAGCGGCGAG 1020
 Db 958 CAACCTCGAGAGATGAGAGGTCTGCAGTGATCAACCCCAATGCGGCGAGCGGCGAG 1017
 QY 1021 CCGCGGTGTGTCAGGAGTGGCTTGCAGACCGGCGAGATGTGTGTTGATGGGCGC 1080
 Db 1018 CCGCGGTGTGTCAGGAGTGGCTTGCAGACCGGCGAGATGTGTGTTGATGGGCGC 1077
 QY 1081 TCGGCGGAAATCTCTGTGGGCGGCGTATCACTAACCAATATGATGCGCTGATCGAC 1140
 Db 1078 ACCGCGGAAATCTCTGTGGGCGGCGTATCACTAACCAATATGATGCGCTGATCGAC 1137
 QY 1141 GAGACCGGCGCTTGTGACGCTGAACGAGATGCGGTCTGAAGAAGCTGAGCGTTGAATAT 1200
 Db 1138 GAAACGCGCATCTGTACCTGTGAACGAGATGCGATCTGAAGAACTGATTTGAATAT 1197
 QY 1201 GAGGTGCGGCGGCTTCTGTGGGTGGGCGGAGCTGTGTCAGCGGCGCACTGAACCGGAG 1260
 Db 1198 GAGGTGCGGCGGCTTCTGTGGGTGGGCGGAGCTGTGTCAGCGGCGCACTGAACCGGAG 1257
 QY 1261 ACCGCGATTACTTCTTGGCGGTCGAACATGCTGTGTAAGATATTAATGCGCGTTGATCA 1320
 Db 1258 AGCGGATCTACTTCTTGGCGGTCGAACATGCTGTGTAAGATATTAATGCGCGTTGATCA 1317
 QY 1321 GAGTTAAGCGCGCTGAGCTTTATACACAGGCGGACCGGCAAACTGCGCGCGGCTTT 1380
 Db 1318 GAATTCACTGAGGACGCTATTAACACAGGCGGACCGGCAAACTGCGCGCGGCTTT 1377
 QY 1381 GAAATATGGGCGCGATCGACGCGATGATATCAGACCGGCGGCACTTGTGTGCGGCG 1440
 Db 1378 GATATATGGGCGCGATCGACGCGATGATATCAGACCGGCGGCACTTGTGTGCGGCG 1437

QY 1441 GAGCGCCCTGGCGGAACTACTCGCCGTTTGTGACAGCGAGCGGCTGTGTTCAAC 1500
 Db 1438 GAGCTGTGCGGCGGAACTATTCCGCCGCTTGTGTGACCGGCGGCGGCTTGTGTTCAAC 1497
 QY 1501 GCGGCGACGACCGCTATTTCGTGCTTCAGCCGAGAAACCGGCGAGACTTTGTGCGAG 1560
 Db 1498 GGTGTGAGGATCTGTTCTTCCGCGCCCTCAGCAAGAAACCGGCGAGACTTGTGCGAG 1557
 QY 1561 GCGCGCTTGGCGAGGTCGCGACGCGGAGGCGATCGATGATGATGATGATGATGATGATGAT 1620
 Db 1558 ACCGCGCTTGGCGAGGTCGCGACGCGGAGGCGATCTTTCAGAGGTTGACGCGATGCA 1617
 QY 1621 TATATCGCATCGGTGCGGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1679
 Db 1618 TATGTGCGCATCGAGGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 1677
 QY 1680 --CGAGGCAATCGATTCGACCTCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 1737
 Db 1678 GCGAGCGAGTCGACTGACCGCGCATCGATGATGATGATGATGATGATGATGATGATGATGAT 1737
 QY 1738 TAA 1740
 Db 1738 TAA 1740
 RESULT 3
 US-09-470-667-3
 ; Sequence 3, Application US/09470667
 ; Patent No. 6730503
 ; GENERAL INFORMATION:
 ; APPLICANT: Asakura, Akira
 ; APPLICANT: Hoshino, Tatsuo
 ; APPLICANT: Ojima, Setsuko
 ; APPLICANT: Shinjoh, Masako
 ; APPLICANT: Tomiyama, No. 6730503;Dumi
 ; TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases
 ; FILE REFERENCE: C38435/109700CON
 ; CURRENT APPLICATION NUMBER: US/09/470,667
 ; CURRENT FILING DATE: 1999-12-22
 ; PRIOR APPLICATION NUMBER: US 08/934,506
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 3
 ; LENGTH: 1737
 ; TYPE: DNA
 ; ORGANISM: Gluconobacter oxydans
 US-09-470-667-3
 Query Match 62.0%; Score 1078.2; DB 4; Length 1737;
 Best Local Similarity 77.1%; Pred. No. 1e-274;
 Matches 1340; Conservative 0; Mismatches 393; Indels 6; Gaps 2;

QY 1 ATGAACCCCAACAGCTGCTTCGACACGCGGCGGCTGTATTTGTTACCGGCGCGC 60
 Db 1 ATGAACCTGACGACCTCTGTGCAAGACGAGCGCGCTGCTGTGCTGTGACCATTC 60
 QY 61 GCATTCCGCGAGATTAACCCCGATTAACGATGAATGCTGTGCGGAAACCGGCGCGTGTGGA 120
 Db 61 GCGCTTGGCC--AACCGCATCAAGATGAATGCTGTGCGGAAACCGGCGCGTGTGGA 117
 QY 121 TGAATTAATGACGCGGCAACCAAGAAATATGCGCATCGGCCCTGACCGATCACT 180
 Db 118 TGAATTAATGACGCGTGAACCAAGAAATATGCGCATCGGCCCTGACCGATTAAC 177
 QY 181 GCCGACACGTTGTGCTGCTTGAACCTGTGTGCGGCGCGGCGATGAGAGGCGGCGCGTA 240
 Db 178 GCAACACACGTCGCGCACTGCAACCTGTGTGCGGCGCGGCGATGAGAGGCGGCGAGATC 237
 QY 241 CAGGTCAAGCGGATGATCACTGATGCGGTGATGATGATGATGATGATGATGATGATGATGATGAT 300
 Db 238 CAATGTGACCGCGCTTGTGCACTGATGATGATGATGATGATGATGATGATGATGATGAT 297

QY	301	CAGGCGCTGGAGTGGCCAAACAGGGGCACTGTAATCTGGAAACACCGCGGCAACTGCGCCG	360
Db	298	CAGGCGCATGAGCGCGCGACCGGCGATCTGATCTGGAAACACCGCGGCAACTGCGCAAC	357
QY	361	GTGCGCATGCTTAAAGCCCAAGCCGACCGCAAGCCGCGGCGTCCGCTTTACGGACGAAC	420
Db	358	ATGCGCACGCTTAAAGCTTTGGTAGCGGACCCCGGCGCATGGCCCTCTATGGACCAAC	417
QY	421	CTCTATTTCAGCTCATGGGACAAACAATCTGATGCGGCTGGATATAGGAGAACGGGCGACGTC	480
Db	418	GTCTATTTTCGTTCTCGTGGGACAACTTGGTGGCTGGACACTCGACCGGCGCAATTC	477
QY	481	GTATTTCGATGTGCAACGTGSAATCGGCGCAAGACGGCTTGACAGTAAACCAACGCGGCGG	540
Db	478	GTATTTCGAGTGTGATCTGGGCGTCAAGACGAAGATA--TGATCTGCAATCTGCGGCGCGG	534
QY	541	ATTGTGCGCAATGGCTGATCTGTCGGGGTTTCACTTCGCAATATTTCCGCTTATGATATGC	600
Db	535	ATTGTGCGCAATGGGCGTATCTGTTGCGGGCTGACCTTCAGTATTTCCGCTTGGGCTGT	594
QY	601	TTTATCTCGGGGCGAGATTCCGCGGAGCGGGTGGAGAGCTGTGGCGCAACACTTTATCCG	660
Db	555	TTTCGTTTGGGCGCAAGACTCGGCGCACGCTGAAAGCTGTGGCGCAACACTTTTATCCG	654
QY	661	CAGCGGCGGAGAGAGGTTGACGAGACTTGGGGCAATGATTTGAGGCGCGCTGGATGAC	720
Db	655	CGCGCGGCGGAGAGAGGTTGATGAGACTTGGGGCAATGATTTGAGGCGCGCTGGATGAC	714
QY	721	GGCGTCTGGGGCTCAATCATCCTATGATCCCGGACGAACCTGTGTTCTATGGCTCGAC	780
Db	715	GGCGTCTGGGGCGCAATCATCCTATGATCCCGTGGCGGCTTTGTCTCATCTGACGACCTCA	774
QY	781	GGCGTGGGCGCCAGCGTCCGAAAACCCAGCGCGGCGACCGCGGCGGCGCTGATGGCAC	840
Db	775	GAAGTTGGCGCTTGGCGGCGAGATTCAGCGCGGCGACGTTGGGCGGCTCATGTATGGCAC	834
QY	841	AACACCCGCTTGGCGGCTGCTGCCGACACGCGGCGAGATTTGCTGGCGCTACCAAGCCGTG	900
Db	835	AACACCCGCTTGGCTGCTGCCGCGCGCGGCGAGATGCTTGGCGTACCAACCTGTG	894
QY	901	CGCGCGGCAACTGGGACCAAGAAATGACAGTTGCGAGATGAGTGTGCGCAACGTCGATGTG	960
Db	895	CGCGCGGCAACTGGGACCAAGAGTGTGAGTGTGCGAGATGAGTGTGTCAAGTGTGCAAGCTTC	954
QY	961	CAACCTCGGCGGAGATGAGGCTTGGCGCTCATCAACCCCAATGGCGCGACGGCGAG	1020
Db	955	CAACCTCGGCTGAGATGAGAGGCTGTGACGCGATCAACCCCGAGCGCGACGGCGAG	1014
QY	1021	CGCGGCTGCTGACGAGTGGCGCTTGGCAAGACCGGCGAGACTGTGCTCTTTGATGCGCGC	1080
Db	1015	CGTCCGCTTGTACCGGCGCTTCCGTGCAAGACGGGCGACATGTGGCACTTTCGACCGGAA	1074
QY	1081	TCGGGCGAATTCCTGTGGCGCGCTGATACCACTACCAATATGATGGCTCGATGCAC	1140
Db	1075	ACCGCGGAATTCCTGTGGCGCGCGGCGACACAGCTATCGAATCTGATGAAAGCGTCAAT	1134
QY	1141	GAGACCGGCTTGTGACGCTGGAACGAGATGGCGGCTGGAAAGAGCTGGAAGTTGAT	1200
Db	1135	CCCGATGGCTGTGTCATGTGACGAGATCTGTGTCTGTGACCGAGCTGAAAGTGGCTTAT	1194
QY	1201	GACGCTGCGCCGACCTTCCCTGGGTGGGCGGCACTGTGCTGACGCGCACTGAACCCGGAC	1260
Db	1195	GAAATCTGCGCGGACCTTCTCTGGGTGGCGGCGGCACTGTGCGCTGAAACCCCGAT	1254
QY	1261	ACCGGCAATTACTTCTTGGCGCTGAAACAATGCTCTCTAGAGATTTATAGGCGGTTGATCAA	1320
Db	1255	ACTGGCACTTATTTATATCCCGCTGAAACAACGCTTGAAGGAGTATAGGCTGTGACCAA	1314
QY	1321	GAGTTTAAAGCGCTGACGCTTATATAACAACGCGGCGACCGCAAACTGCGCGCGGCTTT	1380
Db	1315	GAGTTCAAGTCCGCTGATGTGTATATAACGTCAGGCTGACATAAATCTGCGCGGTTG	1374
QY	1381	GAAAAATATGGCGCGCATGACGCGCATTTGATATACAGACCGGCGGCACTTGTGTGGCGG	1440

Db	1375	GAAACATGGGCGCGATGACGCATCGACATCAGACCGGCGCCACCGTCTGGTCGGCT	1434
Qy	1441	GAGGGCCCTGGGGGAATACTCGCGCGCTTTTGTCAGCGGAGGCGGTGTGTCTCAAC	1500
Db	1435	GAGCGTACGCTCGAATTAAGCGCCTCTCTCTGTCAACGGGCGGGGTCTGTCAAC	1494
Qy	1501	GCGGGACCGACCGCTATTTCCTCGCTCTCAGCCAGCAAAACGGCGAGACTTTGTGGAG	1566
Db	1495	GCGGCGACCGACCGCTACTTCTCGCGCTTCAGCCAAAGACCGGGAGACCTGTGGAG	1554
Qy	1561	GCGCGCTCTTGGACGGTGCAGACCGGGCAGGCGATCAGATTGAGCGGCTGCAA	1620
Db	1555	ACCGGTCTGGAGACTGTGCCTCGGGGTAAACGATTTCTTTGAGATCGAGGGCTGAA	1614
Qy	1621	TATATCGCATCGGAGCGGGCGGTGTGACTATGACAGCATTAAGACGCCCGCTGACC	1680
Db	1615	TATGTGCGCATCGGGCGGGCGGACACAGCTATATGGAGCAACCAACCGGCGCTTGACC	1674
Qy	1661	GAGCGAATCGATTGCACTCGGTGGTAAATCCGATCTATGTCTTTGCACTCCCGAGAT	1739
Db	1675	GAGCGGATCGACTGCAGCCGATCGGACCGCGGAGATATGTCTTTGCTCTCCGACGA	1733

```

RESULT 4
US-09-470-667-2
/ Sequence 2, Application US/09470667
/ Patent No. 6730503
/ GENERAL INFORMATION:
/ APPLICANT: Asakura, Akira
/ APPLICANT: Hoshino, Tatsuo
/ APPLICANT: Ojima, Setsuko
/ APPLICANT: Shinjoh, Masako
/ APPLICANT: Tomiyama, No. 6730503ibumi
/ TITLE OF INVENTION: No. 6730503el Alcohol/Aldehyde Dehydrogenases
/ FILE REFERENCE: C38435/109700CON
/ CURRENT APPLICATION NUMBER: US/09/470,667
/ CURRENT FILING DATE: 1999-12-22
/ PRIOR APPLICATION NUMBER: US 08/934,506
/ PRIOR FILING DATE: 1997-09-19
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 2
/ LENGTH: 1740
/ TYPE: DNA
/ ORGANISM: Gluconobacter oxydans
US-09-470-667-2

```

Query Match	60.6%	Score 1053.6	DB 4	Length 1740
Best Local Similarity	75.7%	Pred. No. 3.2e-268		
Matches 118	Conservative 0	Mismatches 419	Indels 3	Gaps 12

Qy	1	ATGAACCCACAACGCTGCTTGGACCAAGCGCGCGGTGATATGCTTACCGCGCCGCC	60
Dp	1	ATGAAGAGTGTGTTTCTGTCTGTGGACGCGTGGCGGCTTGCAAGCTATAGCTCTTT	60
Qy	61	GCATTGCGCGAGGTAAACCCCGATTACCGATGATCTGTGSGAAACCCGCCCGCTGTGGAA	120
Dp	61	GCGCTTGCTCAAGTACCCCGGTACCCATATATGCTGGSGAAACCCGCCCGTGTGAA	120
Qy	121	TGATTAATCTACGGCGCGCAACCAAGAAATCATGCGCACTCGCCCTTACCAGATCACT	180
Dp	121	TGGATCAGCTACGGTCAGACCAAGAAATCTACGTCACTCGCCCTTACCGATCAAG	180
Qy	181	GCCGCAACGTTGGTCACTTGCCTGTCTGGGCGCGCGGATGGAAGCGGGGCGCTA	240
Dp	181	ACTGAGAACGTGGGCAACTGCACCTGTTGGGGGCGCGGCATGACCGGGCAAAATC	240
Qy	241	CAGGTACGCGCGATGATCATGATGGCGTATATCTTGGCAAAACCCGGGTGATGTATC	300
Dp	241	CAAGTCAGCGCCCTGATCATGACGGCGCTCATGTATCTGGCAAAACCCGGGGAAGCTGTATC	300
Qy	301	CAGGCGCTGATGGGCAACAGGGGATGTATCTGGGAACAACCGCGGCAACTGCCCCC	360


```

; TOPOLOGY: unknown
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Acetobacter alioacetigenes
; STRAIN: MH-24
; PUBLICATION INFORMATION:
; AUTHORS: Tamaki, Toshimi;
; AUTHORS: Fukaya, Masahiro;
; AUTHORS: Takemura, Hiroshi;
; AUTHORS: Tayama, Kenji;
; AUTHORS: Okumura, Hajime;
; AUTHORS: Kawamura, Yoshiva;
; AUTHORS: Nishiyama, Makoto;
; AUTHORS: Horinouchi, Suetaru and
; AUTHORS: Bepko, Teruhiko
; TITLE: Cloning and Sequencing of the Gene Cluster
; TITLE: Encoding Two Subunits of Membrane-Bound
; TITLE: Alcohol Dehydrogenase from Acetobacter
; JOURNAL: Biochimica et Biophysica Acta.
; VOLUME: 1088
; PAGES: 292-300
; DATE: 1991
; US-07-985-458-1

```

```

Query Match      5.1%; Score 88.8; DB 1; Length 2214;
Best Local Similarity 50.8%; Pred. No. 1e-13;
Matches 266; Conservative 0; Mismatches 252; Indels 6; Gaps 2;

```

```

QY 52 GCGCCCGCCGATTCGCGAGTAAACCCGATTAACGATGAACTGTGCGAAACCCGCC 111
DB 97 GCATCCGAGATGACGGGCGAGGCGCCGCGGAGGAGCATATCATGCCGATGATCAG 156
QY 112 GCTGCGAATGATTAATCTACGCGCGGCAACCAAGAAATATGCGCACTGCCCGGAGC 171
DB 157 CCGGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 216
QY 172 CAGATCACTCCGCAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 231
DB 217 CAGATCACTCCGCAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 276
QY 232 GGGGCCCTAAGT--CAAGCCGATGATGATGATGATGATGATGATGATGATGATGAT 288
DB 277 AACCGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 336
QY 289 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 348
DB 337 TGAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 396
QY 349 CAAGTCCGCGCGCTGCGCAAGCTTAAAGCCCAAGCGAC--CGCAAGCGCGCGCTGCC 405
DB 397 CCGGTCGCGCGCAACATGCGCAAGGCGCTGCTGTGACACGCTCAACCGTGGCGGCG 456
QY 406 CTTAAGCGCAAGCGCTTATTCAGTCAATGGAACACCATGATGATGATGATGATGAT 465
DB 457 TACTGGAATGCAAGGCTTATTCAGTCAATGGAACACCATGATGATGATGATGATGAT 516
QY 466 GAGAGCGGCGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 525
DB 517 AAGACCGGCAAGCTGCTGGAAGCTGCAACACCATTCGCGCGCAAGCGAACTGGGCA 576
QY 526 AAGACCGGCGGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
DB 577 CAGCGTCTTAAAGGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620

```

```

RESULT 6
US-09-296-284-22
; Sequence 22, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki

```

```

; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
; TITLE OF INVENTION: and Methods of Use Thereof
; FILE REFERENCE: 1533.0870000
; CURRENT APPLICATION NUMBER: US/09/296,284A
; CURRENT FILING DATE: 1999-04-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 22
; LENGTH: 2163
; TYPE: DNA
; ORGANISM: Glucobacter suboxydans
; US-09-296-284-22

```

```

Query Match      4.9%; Score 85.6; DB 3; Length 2163;
Best Local Similarity 50.3%; Pred. No. 7e-13;
Matches 304; Conservative 0; Mismatches 279; Indels 21; Gaps 3;

```

```

QY 100 GCGAACCAGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 159
DB 28 GCGAACCAGCATCCGCGTGAATGATGATGATGATGATGATGATGATGATGATGATG 87
QY 160 TCGCCCTGACCCAGATCACTGCCGACACGTTGTCAGTTGCACTGTCTGGGCCCCG 219
DB 88 AGCCGCTGATCAGATCAACCAAGACATGCGAATCGAAGCTGCGATGCACTAC 147
QY 220 GGAATGAGGCGGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 276
DB 148 GATCTGATCAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 207
QY 277 CTGGCAACCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 336
DB 208 GCGAACCAGCATCCGCGTGAATGATGATGATGATGATGATGATGATGATGATGATG 267
QY 337 GAAACCGCGCGCACTGCCCGCGCTGCCAAGCTTAAAGCCCAAGCGA--CCGCAAG 393
DB 268 TCTTACATCAAGGTTTCCAGCAACATGCGAAGCGCGCGCTGCTGCTGCTGCTGCT 327
QY 394 GCGCGCGCTGCTTTCAGGCAAGCGCTTATTCAGTCACTGATGATGATGATGATGATG 453
DB 328 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 387
QY 454 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 498
DB 388 GCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 447
QY 499 GATCGGCGGAGGAGCGCTTACCAATGATGATGATGATGATGATGATGATGATGATGAT 558
DB 448 CAGCTGGTCAACGCGCTCTTACAGGTTGAGCGTGTCTCCGATGCGCAAGGCGAAG 507
QY 559 ATGCTGCGGCTTCCAGCTTCAATGATGATGATGATGATGATGATGATGATGATGAT 618
DB 508 GTCAATCACTGCAACGCGCGGTGAGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCG 567
QY 619 TCCGCGAGCGGATGAGAGCTGTGCGCAACCACTTATTCGCGAGCGCGCGGAGAGG 678
DB 568 GCTGAACGGGAGGAGGAGTGGCGCTTTCACCGTTCCGAACCTGCAACCAAGCGG 627
QY 679 GAGC 682
DB 628 GAGC 631

```

```

RESULT 7
US-09-296-284-1
; Sequence 1, Application US/09296284A
; Patent No. 6204040
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: Rhee, Sang-Ki
; APPLICANT: Lee, Eun-Hae
; TITLE OF INVENTION: Glucobacter Suboxydans Sorbitol Dehydrogenase, Genes
; and Methods of Use Thereof

```

```

1 FILE REFERENCE: 1533.0870000
2 CURRENT APPLICATION NUMBER: US/09/296,284A
3 CURRENT FILING DATE: 1999-04-22
4 NUMBER OF SEQ ID NOS: 87
5 SOFTWARE: PatentIn Ver. 2.0
6 SEQ ID NO 1
7 LENGTH: 2265
8 TYPE: DNA
9 ORGANISM: Gluconobacter suboxydans
10 US-09-296-284-1

```

Query Match	4.9%	Score 85.6	DB 3	Length 2265
Best Local Similarity	50.3%	Pred. No. 7.1e-13		
Matches 304; Conservative	0	Mismatches 279	Indels 21	Gaps 3

[illegible]

```

RESULT 8
US-09-296-284-7
;
; Sequence 7, Application US/09296284A
; Patent No. 6204040
;
; GENERAL INFORMATION:
;
; APPLICANT: Choi, Eun-Sung
;
; APPLICANT: Rhee, Sang-Ki
;
; APPLICANT: Lee, Eun-Hae
;
; TITLE OF INVENTION: Glucosyltransferase, Genes
; TITLE OF INVENTION: And Methods of Use Thereof
;
; FILE REFERENCE: 1533.087000
;
; CURRENT APPLICATION NUMBER: US/09/296,284A
;
; CURRENT FILING DATE: 1999-04-22
;

```

```

: NUMBER OF SEQ ID NOS: 87
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 7
: LENGTH: 4830
: TYPE: DNA
: ORGANISM: Gluconobacter suboxydans
:
US-09-256-284-7

```

Query Match	4.9%;	Score 85.6;	DB 3;	Length 4830;
Best Local Similarity	50.3%;	Pred. No. 9.5e-13;		
Matches 304;	Conservative 0;	Mismatches 279;	Indels 21;	Gaps 3;

QY	100	GCGAATCCCGCCCGCTGGTGTGATGGATTACCTACCGCCGCAACCAAGAAATCTATCCGCAC	159
Db	794	GCCGATCAGCATCCGGGTGACTGGATGAGCTATGGCCGGAACCTATTCCGAGCAACGGCTAC	853
QY	160	TCCGCCCTTGACCCAGATCATCTGCCGAAACAGTGTGATTCGAACTCTGGTCCGGCCCGC	219
Db	854	AGCCCGCTGGATCGATCATACCAAGGAAATAGGACGACATCTTAAGCTGGCCATGGCACTAC	913
QY	220	GGGATGAGAGCGGGGGCCGTAACAGTCC--ACGCCGATGATCCATGATGGCGTATGAT	276
Db	914	GATCTGATATCCAAACCGGTGCTCAGGAAGATACGCCGCTGATCTGTTATGGCGTATGATAC	973
QY	277	CTGGCAAAACCCCGGTGATTTGATTCAGAGCGCTGGATGGCAAAACAGGCGATCTGATCTGG	336
Db	974	GCCACCCCAAACTGGAGCAAGATTAAGGCTTTGGATGAGCTACGGGCAAGCTCTGTGG	1033
QY	337	GAAACACCGCCGCCAACTGCCCGCCGTGCCACAGCTAAACGCCCAAGCCGA--CCGCAAG	393
Db	1034	TCTTACGATCCAAAGGTTCCAGAGCAATCGCCACCGCGCTGCTCGATACGCTCAAC	1093
QY	394	CGGGGCGTTCCGCTTTACGGGACAGGCTGATTTCAGCTCATGGAGCAACATCTGATC	453
Db	1094	CGTGGTCAAGCTTACTGGAACGGGAAAGTCTATTTCGGCACTTCGACGCTGGCTGATT	1153
QY	454	GCCTGTGATATGAGACCGGCCAGGTCTGATTCTGATCTC-----GAAAGT	498
Db	1154	GCCCTGTGATGCCAAGACCGGCAAGCTGTCTGGAGCTCTATACGTTTCCCAAGAAACCG	1213
QY	499	GGATCGGAGCGAAGACGGGCTTGAACAGTAAACACACGGGGCCGATTGTTCGCATAGCGCTC	558
Db	1214	CAGCTGGGTACCAACGCGCTCTCTTAACGGTTAGCGGTCTCTCCCGTATGCCAAGGGCAG	1273
QY	559	ATGCTCCGGGTTTCCACTGCGCAATATTCCGCCCTATGAGTGCCTTATCTCGGGGCAAGAT	618
Db	1274	GTCATCTATCGCAACGGCGGTGCAGAGTTTCGGCCGCCGTGGCTTCGTGACGGCTATATGAC	1333
QY	619	TCCCGCAGCGGGTGAAGAGCTGTGGCGCAACCACTTTATCCGCGACCGCGCGCAAGAGGAT	678
Db	1334	GCTGAATGGGAAAGATGGAATGGCTGGCTTTCTTACCGGTTCCGAACCTTGACAAACAGGCG	1393
QY	679	GACG 682	
Db	1394	GACG 1397	

```

; RESULT 9
; US-09-252-991A-15371/C
; Sequence 15371, Application US/09252991A
; Patent No. 6551795
;
; GENERAL INFORMATION:
;
; APPLICANT: Marc J. Rubenfield et al.
;
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;
; FILE REFERENCE: 107196.136
;
; CURRENT APPLICATION NUMBER: US/09/252,991A
;
; CURRENT FILING DATE: 1999-02-18
;
; PRIOR APPLICATION NUMBER: US 60/074,788
;
; PRIOR FILING DATE: 1998-02-18
;
; PRIOR APPLICATION NUMBER: US 60/094,190
;
; PRIOR FILING DATE: 1998-07-27
;
; NUMBER OF SEQ ID NOS: 3142
;

```


SEQ ID NO 15371
LENGTH: 1989
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15371

Query Match 4.4%; Score 77; DB 4; Length 1989;
Best Local Similarity 46.5%; Pred. No. 1.3e-10;
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCCGCACTGAACCCGAGC 1260
DB 696 GAAGTATCGCCGCGCTTCCTCGGCGCAAGAACCTGAACCCGATGCTCAACCCGAGC 637
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACAATGCTGTACGATTTATGCGCTTGATCAA 1320
DB 636 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGAGAACTACTGACCGAGAGGTC 577
QY 1321 GAGTTAGCGCGCTGCAAGCTCTATACACAGCGCCGCAAACTCGCGCGGCTTT 1380
DB 576 AGCTATACGAAGGCAAGCGCTTACCTTGCGATGGGCTTCGAGTCAAGCGCATGACGAC 517
QY 1381 GAAATATGAGCGCGCATGACGCGATGATATACAGACCGGCGCACTTGTGTGCGG 1440
DB 516 GACCAAGCTGCGGACGCTGCGGCGCATGACCGGTCAGCGGCAAGGTGCTGGGAACAC 457
QY 1441 GAGCGCGCTGCGGGAATCTACTGCGCGCTTTGTGCGAGCGCGGTGTGTTCAAC 1500
DB 456 AAGGAACACCTGCGCTGCGGCGCGGAGTGTGCGGCAAGCGCGCAACTGTGTTCAC 397
QY 1501 GCGGAGACCGACCGCTATTTCCGTGCGCTGACGCGAGAAACCGCGGAGACTTGTGCGAG 1560
DB 396 GGCACCGCGGCAAGCGCTTCTCAAGCGCTTGAACGCAAGAGCGGCAAGAGCTGTGAAA 337
QY 1561 GCCGTCTTGGCGAGCGTGCAGCGGCGGCGATGACTACGATTGAGCGGCTGCAA 1620
DB 336 TTCAGACCGCGGACCGCATGCTCTGCGCACGATACCTGGGAAACAGAGCGGCGAGCAG 277
QY 1621 TATATCGCATGAGTGTGCGGCGGCTGACCTATGACGCAATTGAACCGCGCGCGGCG 1680
DB 276 TACCTGCGCGTGAACCGTGTGCGCTACGCGGCGGCGGCTGCGGCGCGGCGGACATGGCC 217
QY 1681 GAGGCATGATTCGACCTGCGTGTGATGATGATGATGATGATGATGATGATGATGATG 1733
DB 216 GACCTGACCGCGCGGCGGCGGCGGCGGCGGCTTCTTGGGATTTCAAGCTGCC 164
```

RESULT 10
US-09-252-991A-15462
Sequence 15462, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074.788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094.190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15462
LENGTH: 2058
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15462

Query Match 4.4%; Score 77; DB 4; Length 2058;
Best Local Similarity 46.5%; Pred. No. 1.3e-10;
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCCGCACTGAACCCGAGC 1260
DB 1495 GAAGTATCGCCGCGCTTCCTCGGCGCAAGAACCTGAACCCGATGCTCAACCCGAGC 1554
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACAATGCTGTACGATTTATGCGCTTGATCAA 1320
DB 1555 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGAGAACTACTGACCGAGAGGTC 1614
QY 1321 GAGTTAGCGCGCTGCAAGCTCTATACACAGCGCCGCAAACTCGCGCGGCTTT 1380
DB 1615 AGCTATACGAAGGCAAGCGCTTACCTTGCGATGGGCTTCGAGTCAAGCGCATGACGAC 1674
QY 1381 GAAATATGAGCGCGCATGACGCGATGATATACAGACCGGCGCACTTGTGTGCGG 1440
DB 1675 GACCAAGCTGCGGACCTGCGGCGCTATGACCGGCTGTGCGGCAAGAGTGTGTGAAAAC 1734
QY 1441 GAGCGCGCTGCGGGAATCTACTGCGCGCTTTGTGCGAGCGCGGTGTGTTCAAC 1500
DB 1735 AAGGAACACCTGCGCTTGTGGGCGGAGTGTGCGGCAACCGCGGCAACCTGTGTTCACC 1794
QY 1501 GCGGAGACCGACCGCTATTTCCGTGCGCTGACCGAGAAACCGCGGAGACTTGTGCGAG 1560
DB 1795 GGCACCGCGGACCGCTTCTCAAGCGCTTGAACGCGGCAAGCGGCAAGAGCTGTGAAA 1854
QY 1561 GCCGTCTTGGCGAGCGTGCAGCGGCGGCGGCGATGACTACGAGTTGACGCGGTCGAA 1620
DB 1855 TTCAGACCGGCAAGCGGCGCATGCTGCGGCGCATGACTGCTGGGAACGCGGAGCAG 1914
QY 1621 TATATCGCATGAGTGTGCGGCGGCTGACCTATGACGCAATTGAACGCGCGGCTGCC 1680
DB 1915 TACCTGCGCGTGAACCGTGTGCGCTACGCGGCGGCGGCTGCGTGTGGGCGGCGGACATGGGCC 1974
QY 1681 GAGGCATGATTCGACCTGCGTGTGATGATGATGATGATGATGATGATGATGATGATG 1733
DB 1975 GACCTGACCGCGCGGCGGCGGCGGCGGCTTCTTGGGATTTCAAGCTGCC 2027
```

RESULT 11
US-09-252-991A-15403
Sequence 15403, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074.788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094.190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 15403
LENGTH: 2637
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15403

Query Match 4.4%; Score 77; DB 4; Length 2637;
Best Local Similarity 46.5%; Pred. No. 1.4e-10;
Matches 248; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

```
QY 1201 GAGCTGCGCCGACCTTCCTGGTGGGCGGACGCTGCTGACCCGCACTGAACCCGAGC 1260
DB 1373 GAAGTATCGCCGCGCTTCCTCGGCGCAAGAACCTGAACCCGATGCTCAACCCGAGC 1432
QY 1261 ACCGCACTTACTCTTCTGCGCTGAACAATGCTGTACGATTTATGCGCTTGATCAA 1320
DB 1433 ACCGCGCTGTTCTACGTCGCGGCAACCACTGGAAGAGAACTACTGACCGAGAGGTC 1492
QY 1321 GAGTTAGCGCGCTGCAAGCTCTATACACAGCGGCGGCGGCTTCTTGGGATTTCAAGCTGCC 1380
```


Db 1493 AGCTATACGAGGAGCGGCTTACTTGGCATGAGGCTTCCGATCAAGCATGTACGAC 1552
 Qy 1381 GAAATATAGGCGCGCATCGACGGATTGATATACGACCGGCGGCGACCTTGTGTGCGCG 1440
 Db 1553 GACCAAGCGGCGGCTTGGCGGCGGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 1612
 Qy 1441 GAGCGCCCTGCGGCGGAACTATCTGCGCCCTTTTGTGCAAGCGGCGGCGGCTGTTCAC 1500
 Db 1613 AAGGAACACCTGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1672
 Qy 1501 GCGGCGGCGGCGGCGGCTTATTTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 Db 1673 GCGACCGGCGGCGGCGGCTTATTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1732
 Qy 1561 GCGCGCTTGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
 Db 1733 TTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1792
 Qy 1621 TATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
 Db 1793 TACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1852
 Qy 1681 GAGCGCATCGATTCGACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1733
 Db 1853 GACCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1905

RESULT 12
 US-09-252-991A-15512
 ; Sequence 15512, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 15512
 ; LENGTH: 510
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-15512

Query Match 4.2%; Score 72.6; DB 4; Length 510;
 Best Local Similarity 48.0%; Pred. No. 1.1e-09;
 Matches 207; Conservativity 0; Mismatches 224; Indels 0; Gaps 0;
 Qy 1201 GACGTCTGCGGCGGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
 Db 78 GAAGTATCGCGGCGGCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 137
 Qy 1261 ACCGCGCTTACTTCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
 Db 138 ACCGCGCTTACTTCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1397
 Qy 1321 GAGTTAGCGCGCTGCACTGTATTAACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
 Db 198 AGCATATCGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 257
 Qy 1381 GAAATATAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
 Db 258 GACCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 317
 Qy 1441 GAGCGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
 Db 318 AAGGAACACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 377

Qy 1501 GCGGCGGCGGCGGCGGCTTATTTCCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
 Db 378 GCGACCGGCGGCGGCGGCTTATTTCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 437
 Qy 1561 GCGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
 Db 438 TTCGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 497
 Qy 1621 TATATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
 Db 498 TACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 508

RESULT 13
 US-09-489-039A-5265
 ; Sequence 5265, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 5265
 ; LENGTH: 1809
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-5265

Query Match 3.1%; Score 53.8; DB 4; Length 1809;
 Best Local Similarity 60.5%; Pred. No. 0.00016;
 Matches 107; Conservativity 0; Mismatches 67; Indels 3; Gaps 1;
 Qy 1472 TGTGACGCGGCGGCGGCTTGTGTTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1528
 Db 1565 TGTGACGCGGCGGCGGCTTGTGTTCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1624
 Qy 1529 TCAGCGAAGAAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1588
 Db 1625 ACAATATAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1684
 Qy 1589 AGCGATGAGTACGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1645
 Db 1685 CTCGATGAGTACGAGTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1741

RESULT 14
 US-08-476-519-10
 ; Sequence 10, Application US/08476519
 ; Patent No. 5750876
 ; GENERAL INFORMATION:
 ; APPLICANT: Barry, Gerard F.
 ; APPLICANT: Kishore, Ganesh M.
 ; TITLE OF INVENTION: No. 5750876el Isoamylase Gene, Compositions
 ; TITLE OF INVENTION: Containing It and Methods of Using Isoamylases
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Grace L. Bonner, Monsanto Company, B34F
 ; STREET: 700 Chesterfield Parkway No. 5750876th
 ; CITY: St. Louis
 ; STATE: Missouri
 ; COUNTRY: USA
 ; ZIP: 63198
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: IBM PC compatible
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 14:45:05 ; Search time 933 Seconds

(without alignment)
10077.497 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1 atgaccccccaacgctgct.....ttcttgcaactgcgcagcagaa 1740

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3627888 seqs, 270181610 residues

Total number of hits satisfying chosen parameters: 7255776

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCR_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCRUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1206.2	69.3	1743	US-10-162-713-4	Sequence 4, Appl1
2	1206.2	69.3	2281	US-10-162-713-11	Sequence 11, Appl1
3	1071.2	61.6	1737	US-10-162-713-1	Sequence 1, Appl1
4	1071.2	61.6	2519	US-10-162-713-9	Sequence 9, Appl1
5	1052	60.5	1740	US-10-162-713-3	Sequence 3, Appl1
6	1052	60.5	3200	US-10-162-713-10	Sequence 10, Appl1
7	152.2	8.7	2151	US-10-369-493-34527	Sequence 34527, A
8	99.4	5.7	1731	US-10-369-493-45426	Sequence 45426, A
9	83.8	4.8	1677	US-10-369-493-28294	Sequence 28294, A
10	76.6	4.4	1701	US-10-369-493-31051	Sequence 31051, A
11	76.6	4.4	2076	US-10-369-493-44211	Sequence 44211, A
12	74.8	4.3	1596	US-10-369-493-31374	Sequence 31374, A

13	74.8	4.3	2051	US-10-369-493-28615	Sequence 28615, A
14	74	4.3	1869	US-10-369-493-31733	Sequence 31733, A
15	68.8	4.0	2120	US-10-369-493-34525	Sequence 34525, A
16	65.8	3.8	2412	US-10-369-493-28508	Sequence 28508, A
17	65.8	3.8	2412	US-10-369-493-31267	Sequence 31267, A
18	62.8	3.6	1776	US-10-369-493-31375	Sequence 31375, A
19	62.8	3.6	1803	US-10-369-493-28617	Sequence 28617, A
20	61	3.5	2331	US-10-369-493-31725	Sequence 31725, A
21	57	3.3	2376	US-10-369-493-28565	Sequence 28565, A
22	57	3.3	2376	US-10-369-493-31324	Sequence 31324, A
23	53.6	3.1	2394	US-10-369-493-39100	Sequence 39100, A
24	53.6	3.1	2394	US-10-369-493-39468	Sequence 39468, A
25	53.6	3.1	2394	US-10-369-493-39850	Sequence 39850, A
26	53.6	3.1	4412	US-09-927-827-43	Sequence 43, Appl1
27	50	2.9	1776	US-10-369-493-39656	Sequence 39656, A
28	50	2.9	1776	US-10-369-493-40018	Sequence 40018, A
29	50	2.9	1779	US-10-369-493-39280	Sequence 39280, A
30	49.4	2.8	2400	US-10-369-493-37751	Sequence 37751, A
31	47	2.7	7155	US-10-329-079-14	Sequence 14, Appl1
32	47	2.7	37360	US-10-329-079-6	Sequence 6, Appl1
33	46.6	2.7	536	US-10-338-110-119	Sequence 119, App
34	46	2.6	2391	US-10-369-493-24390	Sequence 24390, A
35	45.8	2.6	1401	US-09-758-759-106	Sequence 106, App
36	45.8	2.6	109519	US-09-758-759-1	Sequence 1, Appl1
37	45	2.6	2358	US-10-369-493-35570	Sequence 35570, A
38	44.6	2.6	1899	US-10-156-761-5926	Sequence 5926, Ap
39	44.6	2.6	9025608	US-10-156-761-1	Sequence 1, Appl1
40	44.4	2.6	851	US-10-425-115-12182	Sequence 12182, A
41	44.4	2.6	893	US-10-425-115-12183	Sequence 12183, A
42	44.2	2.5	1860	US-10-369-493-40752	Sequence 40752, A
43	43.6	2.5	894	US-10-282-1224-25434	Sequence 25434, A
44	43.6	2.5	1338	US-10-369-493-31938	Sequence 31938, A
45	43.4	2.5	1049	US-10-123-155-358	Sequence 358, App

ALIGNMENTS

```
RESULT 1
US-10-162-713-4
; Sequence 4, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Sul-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: Sorbital Dehydrogenases of Ketogulonigenium spp., Genes and Methoc
; FILE REFERENCE: 1533.3740000
; CURRENT APPLICATION NUMBER: US/10/162,713
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1743
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-4
Query Match 69.3%; Score 1206.2; DB 15; Length 1743;
Best Local Similarity 80.9%; Pred. No. 0;
Matches 1406; Conservative 0; Mismatches 333; Indels 0; Gaps 0;
QY 1 ATGACCCCAACACCGTCTGCGACCGCGGCGCGGCTATTGCTTACCGCGCCGCC 60
Db 1 ATGACCCCAACACCGTCTGCGACCGCGGCGCGGCTATTGCTTACCGCGCCGCC 60
QY 61 GCATTGCGCGACGTAACCCCGATTACGATGAATCTGCGGACCCCGCGCTGTGTA 120
```

Dp	61	GCCTTGGGAGGCTGACCCCCATCAACGATGAACTGTGGCCAAACCCGCGACGGGGAG	120
Qy	121	TGATTAACTACGGCCGGCAACCAAGAAACTATCCGCACTGGCCCTTGACCCAGATACT	180
Dp	121	TGATTCAACTACGGCCGGGATCTGGGAAAACTACCCCACTGGCCCTTGGAACAATTAACG	180
Qy	181	GCCACAAAGCTGGATCGTTGCAACTGGACTGAGGGCCGCGGGGATGGAAGCGGGGGCGTA	240
Dp	181	ACCGACAAAGCTGGCCAGCTGCAAGCTGTGCTTGGGGCCGGGCAATGGAAGCGGGCGCTG	240
Qy	241	CAGGTCAAGCCGATGATTCATGATGAGCGTGATATCTTGGCAAAACCCCGGTGATGATC	300
Dp	241	CAGGTCAACCCCGATGATTCACAGCGCGCTGATGTCTGGCCAAACCCCGGCGAGTATC	300
Qy	301	CAGGCGGTGATTCGGCAACAGCGGATCTGATCTTGGGAACACCCCGCGCAATCCGCCCC	360
Dp	301	CAGGCAATCAACGCAAAACCGCGACTGATGTGGGAACACCGCGCAACTCCGCCCC	360
Qy	361	GTCGCAACGCTAAACGCCCAAGCGCAAGCGCGCGCTTGACCGGCAACAGC	420
Dp	361	GTTGCTCTGTTGAACGGCCAAAGCGACGTTAAACGGCGGTGGCCCTTATGGACCAAC	420
Qy	421	CTCATTTTCAAGCTCATGGGAACAACCTGATTCGCGCTTGATATGGAACGGGCAAGTC	480
Dp	421	CTCATTTTCAACCTGTGGGAACAACCTTGTGCACTGGAACATGGGCAACGGGCAAGTC	480
Qy	481	GTATTGCAATGTCAGTGGATCGGGCGGAAGACGCTTGAACAGTAACAACAGGGGCG	540
Dp	481	GTCCTTGAATGTGACAGCGCGCTTCGGCGATGACGGGCTGACACACACACAGGCGCG	540
Qy	541	ATTGTGCCAAATGAGCCTCATTCGTCCGCGGTTCACTTCCCAATATTGCCCATGATGC	600
Dp	541	ATTGTCCAAACGGCCTCATTCGTCCGCGCTGCACTTCCCAATCTCGCCTTACGCTGC	600
Qy	601	TTTATCTCGGGGACGATTCGCGGAGCGGGTAGAGGCTGTGGGCAACACTTATTCGG	660
Dp	601	TTTGTCTCGGGTACCACTCGGCGAGCGCGGCAAGACTGTGGGCAACTTCAATCCG	660
Qy	661	CAGCGGGCGAAGAGGTGACGACTTGGGGCAATGATTTGAGGGCGCGTGGATGAC	720
Dp	661	CAAGCGGGCGAAGAGGCGACGACGACTTGGGGCAACACTTGATGCGCGTGGATGAC	720
Qy	721	GGGCTCTGGGGTCAATCACTTATGATCCCGTGAAGAACTTGTGTTCTATGGCTGAC	780
Dp	721	GGGCTCTGGGGCCAGCTGACTATGACCCCGTCAACAATCTGTGTGCACTGCGCTGACC	780
Qy	781	GGCGTGGGCCCAAGCTGCCAAACCAAGCGCGGCAAGCGCGGCGCAAGCTGTATGGACC	840
Dp	781	GGCGTGGGCCCCCGCACTCGGAACCCAGCGCGGCAACCCCGGCGCAAGCTTATGGGACC	840
Qy	841	AACACCCGCTTTGGGTCGTCCCAACAGCGGCGAATTTCTTGGCTCAACGACCCTG	900
Dp	841	AATACCCGCTTTGCCCTGTGCCCCCAACAGGCGGAATTCGTGGCGCCACCAACCTTG	900
Qy	901	CCGGCGGACAACCTGGGACCAAGATGCACTTCCAGATGATGGTCCGCAACGTCAATGG	960
Dp	901	CCCCGGAACACTGGGACCAAGATGCACTTCCAGATGATGGTCCGCAACGTCAACGTG	960
Qy	961	CAACCTTCGCGCGAGATGGAAGGCTTGGCGGCATCAACCCCAATGCGAGACGCGGAG	1020
Dp	961	CAACCCGCTTCGCGACATGGAAGGCGTTGAGGCGATCAACCCCAACGCAACATTCGCGAG	1020
Qy	1021	CGCGGTGTGTCGACGGGTGCGCTTGCAAAAGCCGGAACGAATGTGTGATGGGGCC	1080
Dp	1021	CGTGGGTGTGACGGGCAATCCCTTGCAAAACCCGGAACCAATGTGTGATGGGGCC	1080
Qy	1081	TCGGGCGAAATTCCTGTGGCGCGTATACCACTACCAATATGATGCGCTGATGCAC	1140
Dp	1081	ACCGGCGAAATTCCTGTGGGCAAGGCAACCACTTACGAACCTGATGCGCTGATGCAC	1140
Qy	1141	GAACTCGGCTTGTGACGCTGAACGAGATGCGGTGCTGAAGAAGCTGCACTGAAAT	1200

QY	DB	Sequence	Score	DB	Length	Matches	Mismatches	Indels	Gaps
QY	1141	GAAACCGGCGCTGATCGATCGATGAACGAAGA	69.3%	1200	2281	0	0	0	0
QY	1201	GACGCTTGCCCGACACTTCCCTGGGTGGGCGGGA	80.9%	1260	2281	0	0	0	0
DB	1201	GACACTTGCCCGACACTTCCCTGGGTGGGCGGGA		1260	2281	0	0	0	0
QY	1261	ACCGGCAATTTACTTCTTGCCGCTGAACAAATG		1320	2281	0	0	0	0
DB	1261	AGCGGCAATTTACTTCTTGCCGCTGAACAAATG		1320	2281	0	0	0	0
QY	1321	GAGTTTACGCGCTTCGACGCTCTATACACCAAG		1380	2281	0	0	0	0
DB	1321	GAAATTCGCGGCTTGAATGATGAACAAATG		1380	2281	0	0	0	0
QY	1381	GAAATTCGCGGCTTGAATGATGAACAAATG		1440	2281	0	0	0	0
DB	1381	GAAATTCGCGGCTTGAATGATGAACAAATG		1440	2281	0	0	0	0
QY	1441	GAGGCGCCCTGGGCGGAATCTACGCGCTTTT		1500	2281	0	0	0	0
DB	1441	GAGGCGCCCTGGGCGGAATCTACGCGCTTTT		1500	2281	0	0	0	0
QY	1501	GAGGCGGACCGACCGGCTATTCCTGCGCCCT		1560	2281	0	0	0	0
DB	1501	GAGGCGGACCGACCGGCTATTCCTGCGCCCT		1560	2281	0	0	0	0
QY	1561	GCCCCCTTTCGACGCGTCCGACCGGCGGCG		1620	2281	0	0	0	0
DB	1561	ACCGGCTTTCGACGCGTCCGACCGGCGGCG		1620	2281	0	0	0	0
QY	1621	TATATCGCATTCGATCGGCGGCGGCGGCGG		1680	2281	0	0	0	0
DB	1621	TATATCGCATTCGATCGGCGGCGGCGGCGG		1680	2281	0	0	0	0
QY	1681	GAGGCAATTCGATTCGATTCGATTCGATTCG		1739	2281	0	0	0	0
DB	1681	GAGGCAATTCGATTCGATTCGATTCGATTCG		1739	2281	0	0	0	0

QY 61 GCATTGCGGAGTAAACCCGATTAACCGATGAAGTCTGGCGAAACCCGCCGCTGTGAA 120
DB 590 GCGTTGGCGGAGTCAACCCCACTACCGATGAAGTCTGGCGAAACCCGCCGCGGCGAG 649
QY 121 TGGATTAAGTACGCGCGGCAACCAAGAAAATATCGGCACTCGCCCTTGAACCAATCACT 180
DB 650 TGGATTAAGTACGCGCGGCAATCAGAAAATATCGGCACTCGCCCTTGAACCAATCACT 709
QY 181 GCGCAAGTGTGTGATGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
DB 710 ACCGACAAAGT 769
QY 241 CAGGTACCGCGGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 770 CAGGTACCGCGGATGATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 829
QY 301 CAGGCGCTGGAATGCGCAACAGGCGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 830 CAGGCGCTGGAATGCGCAACAGGCGATGATGATGATGATGATGATGATGATGATGATGAT 889
QY 361 GTCGCGACGCTAAACGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 890 GTTGCTCGCTGAACG 949
QY 421 CTCTATTCACTCATGGAACCACTCTGATGCGCGCTGGAATGGAAGCGGCGCGCGCGCG 480
DB 950 CTCTATTCACTCATGGAACCACTCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009
QY 481 GTATTGATGATGGAAGT 540
DB 1010 GTTTTGAATGTGAAGCG 1069
QY 541 ATTGTGCGCAATGCGCTCATGCTGCGCGGTTTCACTGCGCAATTTGCGCTTATGATGATG 600
DB 1070 ATTGTGCGCAATGCGCTCATGCTGCGCGGTTTCACTGCGCAATTTGCGCTTATGATGATG 1129
QY 601 TTTTATCTCGGCGGACGATTCGCGGACGCGGATGGAAGTGTGTGTGTGTGTGTGTGTGT 660
DB 1130 TTTTATCTCGGCGGACGATTCGCGGACGCGGATGGAAGTGTGTGTGTGTGTGTGTGTGT 1189
QY 661 CAGCGCGGCGAAGAGGATGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGAC 720
DB 1190 CAGCGCGGCGAAGAGGATGACGAGACTTGGGCGAATGATTTGAGGCGCGCTGATGAC 1249
QY 721 GCGGCTCGGCGGATGATCACTTATGATCCCGTGAACCACTTGTGTGTGTGTGTGTGTGT 780
DB 1250 GCGGCTCGGCGGATGATCACTTATGATCCCGTGAACCACTTGTGTGTGTGTGTGTGTGT 1309
QY 781 GCGGCGGCGGACGATTCGCGAACCAGCGCGGACGCGCGGCGCGCGCGCGCGCGCGCGCG 840
DB 1310 GCGGCGGCGGACGATTCGCGAACCAGCGCGGACGCGCGGCGCGCGCGCGCGCGCGCGCG 1369
QY 841 AACACCGGCTTGTGCGGTGCTGCGGACGCGGCGGAGATTTGTGCGGTGACGAGCTTG 900
DB 1370 AACACCGGCTTGTGCGGTGCTGCGGACGCGGCGGAGATTTGTGCGGTGACGAGCTTG 1429
QY 901 CCGCGGCGACATCGGCGGACCAAGATGAGTGTGAGATGATGATGATGATGATGATGATGAT 960
DB 1430 CCGCGGCGACATCGGCGGACCAAGATGAGTGTGAGATGATGATGATGATGATGATGATGAT 1489
QY 961 CAAACCTCGGCGGAGATGAGAGGATGCGGCGCATCAACCCCAATGCGGCGAGCGGCGAG 1020
DB 1490 CAAACCTCGGCGGAGATGAGAGGATGCGGCGCATCAACCCCAATGCGGCGAGCGGCGAG 1549
QY 1021 CGCGGTGTGAGCGGATGCGCTTGAAGACCGGCGCAATGTGTGTGTGTGTGTGTGTGTGT 1080
DB 1550 CGCGGTGTGAGCGGATGCGCTTGAAGACCGGCGCAATGTGTGTGTGTGTGTGTGTGTGT 1609
QY 1081 TCGGCGCAATTCCTGTGCGCGGATGATCACTACCAATGATGATGATGATGATGATGATGAT 1140
DB 1610 TCGGCGCAATTCCTGTGCGCGGATGATCACTACCAATGATGATGATGATGATGATGATGAT 1669
QY 1141 GAGACCGGCTTGTGAGCGTGAACGAGAGATGCGGTGTGAAGAGCTGACGATGATAT 1200

DB 1670 GAAACCGGCTGTGACAGTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1729
QY 1201 GAGCTTGGCGGACCTTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1260
DB 1730 GAGCTTGGCGGACCTTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1789
QY 1261 ACCGCAATTTACTTCTTCCGCTGAACCAATGCGCTGATGATGATGATGATGATGATGATGAT 1320
DB 1790 AGCGGCAATTTACTTCTTCCGCTGAACCAATGCGCTGATGATGATGATGATGATGATGATGAT 1849
QY 1321 GAGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1380
DB 1850 GAGTTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1909
QY 1381 GAAATATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1440
DB 1910 GAAATATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1969
QY 1441 GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
DB 1970 GAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2029
QY 1501 GCG 1560
DB 2030 GCG 2089
QY 1561 GCGCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
DB 2090 ACCGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2149
QY 1621 TATATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1680
DB 2150 TATATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2209
QY 1681 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739
DB 2210 GAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2268

RESULT 3
US-10-162-713-1
; Sequence 1, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; APPLICANT: Yum, De-Young
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method
; FILE REFERENCE: Use Thereof
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-1
Query Match 61.6%; Score 1071.2; DB 15; Length 1737;
Best Local Similarity 76.4%; Fred. No. 1.2e-308;
Matches 1329; Conservative 0; Mismatches 408; Indels 3; Gaps 1;
QY 1 ATGAACCCCAACAGCTGCTTGCACACGCGCGCGGCTGATTTGCTTACCGCGCGCGCG 60
DB 1 ATGAATGGAATTCGTTCTTGTGCGAAGCGTGTGCGGCTTGTGATTTGCTGCTGCGCC 60

```

Qy 61 GCATTCCGCGAGTAACCCCGATTACCGATGAATCTGTGGCAACCCCGCCCTGTGTGA 120
Db 61 GCATTTCGCGATGAGCCCGGTCAACGAGCTGTGTGCAAAACCCCGCCCGCGGAA 120
Qy 121 TGGATTACTAGCGCGCAACAAGAAAATTATCGCACTCGCCCTGACCCGATCACT 180
Db 121 TGGATTAGCTATGGCCGCAACAAGAAAATTATCGCACTCGCCCTGACCCGATCACT 180
Qy 181 GCCGACAACTGTGTGCTGCACTGTGTGTGGGCGCGGGAGTGAAGCGGGGCGCTGA 240
Db 181 CCGGACAACTGTGTGCGCGCTGTGTGTGGGCGCGGGAGTGAAGCGCGCGCTGTGA 240
Qy 241 CAGGTCAAGCGGATGATCACTGAATGGGTGATGATGTATCTGGCAAAACCCCGGTGATGATC 300
Db 241 CAGGTCAAGCGGATGATCACTGAATGGGTGATGATGTATCTGGCAAAACCCCGGTGATGATC 300
Qy 301 CAGGCGCTGAGTGCAGCAACAGCGATCTGTATCTGGCAACCCCGCGCAACTGCGCGCC 360
Db 301 CAGGCGCTGAGTGCAGCAACAGCGATCTGTATCTGGCAACCCCGCGCAACTGCGCGCC 360
Qy 361 GTGGCAGCGCTAAACCGCCAGCGCAACCGACCGCGCGCTGCGCTTTACCGCAAGAC 420
Db 361 ACTCGACGCTCAGCTGTGGGGAGTGCAGCGCGCATGCGCTTTATGCGCAACT 420
Qy 421 CTCTATTCACTCATGTGGAACAACAATCTGATCGCTGATATGAGACGCGCGCAGTTC 480
Db 421 GTCTACTTCTGTGTGGGCAACAACAATGTGTGCGCTGATCTGCGCGCGCAAGTC 480
Qy 481 GTATTGATGTGCAAGTGTGATCGGGGAAAGACGCGCTTGAACCTAAACACAGGGGCGC 540
Db 481 GTCTTGAAGCTGCAACGCGCGCAAGGCAAGCGAGCGGCTGTG--AACTGTGTGCGCGCC 537
Qy 541 ATTGTGCGCAATGAGCGTCACTGTGCGGGGTTCACCTGCAATATTCGCGCTATGATGC 600
Db 538 ATTGTGCGCAATGAGCGTCACTGTGCGGGGTTCACCTGCAATATTCGCGCTATGATGC 597
Qy 601 TTTATCTCGGGGCAAGATTCGCGACGCGGTGAAGACCTGTGGCGCAACCACTTTATCCG 660
Db 598 TTTGTGTGCGGCGCATATGCAAGCAAGCGGCAAGCACTGTGGCGCAACTTTATCCG 657
Qy 661 CAGCGGGGCAAGAGGGTGAAGCACTGTGGGCAATTTTGGAGGCGCGGTGATGAC 720
Db 658 CAGGCAAGTGAAGAGGTGAAGCAACTGTGGGCAATTTTGAAGGCGCGGTGATGAC 717
Qy 721 GAGGTCTGGGGTGAATCACTATGATCCGTGACGAACCTTGTGTCTATGAGCTGAC 780
Db 718 GAGGTCTGGGGTGAATCACTATGACCCCACTAATTTGTGTCTATGAGATGCTG 777
Qy 781 GGGGTGGGCGCCAGCGTCCGAAACCCAGCGCGGCAACCGCGCGCGCTGATGAC 840
Db 778 GCGGTAGGCGCGCGCATCCAGGTTCAAGCGCGGCAACCGCGCGCGCTTTACGCGAC 837
Qy 841 AACACCGGCTTTGCGGTGCGTCCGCAACAGCGGCGAGATTTGTGTGCGTCAACGACCTG 900
Db 838 AACACCGGCTTTGCGGTGCGTCCGCAACAGCGGCGAGATTTGTGTGCGTCAACGACCTG 897
Qy 901 CCGCGGCAACACTGGAACCAAGATCAAGTGTGAGATATGTGTGCGCAACGTCATGTG 960
Db 898 CCGCGGCAACACTGGAACCAAGATCAAGTGTGAGATATGTGTGCGCAACGTCATGTG 957
Qy 961 CAACCTCGGCGAGATGAGAGGATCTGCGCGGCGCATCAACCCCAATGCGCGCAAGCGAG 1020
Db 958 CAGCGCGGCTGCGCAATGAGAGGCGCGCATCAACCCCAATGCGCGCAAGCGAG 1017
Qy 1021 CCGCGGTGTGCTGAGCGGTGCGCTTGAAGACCGGCAAGATGTGTGCTTGTATGCGGC 1080
Db 1018 CCGCGGTGTGCTGAGCGGTGCGCTTGAAGACCGGCAAGATGTGTGCTTGTATGCGGC 1077
Qy 1081 TCGCGGCAATTCCTGTGGGCGGTGATACCAACTATACCAATATGATGCGCTGATGAC 1140
Db 1078 ACGGGGAAATTCCTGTGGGCGGTGATACCAACTATACCAATATGATGCTGATGAC 1137

```

```

Qy 1141 GAGACCGGCTCTTGTGACGCGTGAACGAGATGCGGTGCTGAAAGCTGGAATAT 1200
Db 1138 GAAACCGGTCTGTGTACGCGTGAATGAGATATCTCTAAAGATGTGACACCGACTAC 1197
Qy 1201 GACGCTGCGCGACCTTCTGTGGTGGCGGCACTGTGTGTCAGCGGCACTGAACCCGAC 1260
Db 1198 CGCATTTGCCGCAATCTTGTGGTGAAGCGGATGCGCGGTGCGGATCTTGAACCCGAT 1257
Qy 1261 ACCGCACTTATCTTCTGCGCGCTGAACAAATGCTGTCAATATATATGCGCGTATCA 1320
Db 1258 AGCGCATCTATCTTCTTCTTCCCTGAACAAAGCGCTGTGCGGAGTGTGATCA 1317
Qy 1321 GAGTTTACGCGCTGACAGTCTATTAACAGCGCGCAACCCCAACTCGCGCGCGCTTT 1380
Db 1318 GAGTTTACGCGCAATGAGAGCTTAAACAGCGCGCACTTACCTGCTGTGCGCGGAAAA 1377
Qy 1361 GAAATATGCGCGCATGACGCGATGATATGACACCGCGCGCACTTGTGTGCGG 1440
Db 1378 GAAATATGCGCGCATGACGCGATGATATGACACCGCGCGCACTTGTGTGCGG 1437
Qy 1441 GAGGCGCGCTGCGGAACTACTGCGCGCTTTGTGTGACGCGCGGCGGTGTGTCAAC 1500
Db 1438 GAGGTCTGTGTGAACTACTGCGCGCTTGTGTGACGCGCGGCGGTGTGTCAAC 1497
Qy 1501 GCGGGAACGACGCGTATTTCCGTGCGCTGACCGCAAGAAACCGCGAGACTTGTGCGAG 1560
Db 1498 GCGGGAACGACGCGTATTTCCGTGCGCTGACCGCAAGAAACCGCGAGACTTGTGCGAG 1557
Qy 1561 GCGCGTCTGTGAGCGTGTGCGAGCGGCGAGCGCATGAGTGAAGTGTGAAGCGCGGCA 1620
Db 1558 ACCGCTGTGAGTGTGCGAGCGGCGAGCGCATGAGTGAAGTGTGAAGCGCGGCA 1617
Qy 1621 TATATGCGCATGCGTGTGCGGCGGTGTGACCTATGACGCAATGAAACCGCGCGTGGCC 1680
Db 1618 TATATGCGCATGCGGCGGCGGTGTGACCTATGACGCAATGAAACCGCAATTCGCG 1677
Qy 1681 GAGGCAATGCAATGACCTGTGTGTGTGATGCGATCTATGTCTTGTGACCTGCGGACTTA 1740
Db 1678 GCGACCAATGCAATGACCTGTGTGTGTGATGCGGCAAGCGGTGTGCTTGTGCGCAATTA 1737

```

```

RESULT 4
US-10-162-713-9
; Sequence 9, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eui-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method
; FILE REFERENCE: 1533 3740000
; CURRENT APPLICATION NUMBER: US/10/162,713
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 2519
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-9
Query Match 61.6%; Score 1071.2; DB 15; Length 2519;
Best Local Similarity 76.4%; Pred. No. 1.3e-308;
Matches 1329; Conservative 0; Mismatches 408; Indels 3; Gaps 1;
Qy 1 ATGAAACCCCAACGCGCTTGTGACGCGGCGCGGTGCTATGCTTACCGGCGCCGCC 60
|||||

```

Db 750 ATGAATTCGAAATTCGTTGCTTCGGAAGGCTTGCCTGCTGATCTTTGCTGTCGCC 809
Qy 61 GCATTGCGGAGGTAACCCCGATTACCGATGAATCTGCTGGCAACCCCGCTGTGTGA 120
Db 810 GCAATTCGCGATGTAGCCCGCTACCGAGAGAGCTGTAGCAAAACCCCGCCGCGGAA 869
Qy 121 TGAATTAACAGCGCGCAACAAGAAATCTAGCCACTGCGCCCTGACCAAGTCACT 180
Db 870 TGAATCACTATGCGCGCAACAAGAAATCTAGCCACTGCGCCCTGACCAAGTCACT 929
Qy 181 GCCGCAACGTTGCTGATGCACTGCTGAGCCCGCGGATGAGAGCGCGGCGTA 240
Db 930 CCGGACAACTGCGCGAGCTGAGCTGCTGAGCGCGGATGAGAGCGCGGCGTA 989
Qy 241 CAGGTACCGCGATGATCCATGATGAGGATGATGATGAGCAACCCCGGATGATGATC 300
Db 990 CAGGTACCGCGATGATCCATGATGAGGATGATGATGAGCAACCCCGGATGATGATC 1049
Qy 301 CAGGCGCTGAGTACGCAACAGGCGATCTGATCTGGAAACCCCGCGCACTGCGCGC 360
Db 1050 CAGGCGATGACGCAAAACCGGATGATCTGAGAAACACCGCGCGCACTGCGCGAG 1109
Qy 361 GTGCGCAAGTAAACGCGCAAGCGACCGGACCGCGCGCTGCGCTTTACGCAAGAC 420
Db 1110 ACCTCGACGCTGAGCTGCGGAGATCGCAACCGCGCATGCGCTTTATGACCAAT 1169
Qy 421 CTCTATTACGCTCATAGGCAACCACTGATCGCTGATATGAGAAGCGCGCAAGTTC 480
Db 1170 GTCTACTGCTGCTGCGGCAACCACTGATCGCTGATATGAGAAGCGCGCAAGTTC 1229
Qy 481 GTATTGATTCGAACTGTGATCGGCGAAAGACGCTTGAACCACTAACCAACGCGCGC 540
Db 1230 GTCTGACGCTGACCGCGCGCAAGGCGAAGCGCGGATGCG---AACTGTCGCGCCC 1286
Qy 541 ATTGCGCAATGAGCGCTCATGCTGCGGAGTTCCACTGCAATATTCGCTATGATGAC 600
Db 1287 ATTGCGCAATGAGCGCTCATGCTGCGGAGTTCCACTGCAATATTCGCTATGATGAC 1346
Qy 601 TTATCTCGGAGGCAAGATTCGCGACGAGTGAAGAGCTGTGGGCAACCACTTTATCCG 660
Db 1347 TTATGTCGAGGCAATGATGCAACGAGCGGAACTGTGGGCAACCACTTTATCCG 1406
Qy 661 CAGCGCGCAAGAGGCTGACGAGCTTGGGCAATGATTCAGAGCGGCTGATGATC 720
Db 1407 CAGCGAGTAAAGGCTGACGAACTGTGGGCAATGATTCAGAGCGGCTGATGATC 1466
Qy 721 GGGGCTGGGAGTCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 780
Db 1467 GGGGCTGGGAGTCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 1526
Qy 781 GGGGCTGGGAGTCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 840
Db 1527 GGGGCTGGGAGTCACTTATGATCCCGTGAAGCACTTGTGTTATGAGCTGAC 1586
Qy 841 AACACCGCTTGGGAGTCCCGACACGCGGAGATTTGCTGAGCTCACGACCTG 900
Db 1587 AACACCGCTTGGGAGTCCCGACACGCGGAGATTTGCTGAGCTCACGACCTG 1646
Qy 901 CCGGCGACAACTGGGACCAAGATGAGCTTGAAGTGTGCGCAACGTGATGTG 960
Db 1647 CCGGCGACAACTGGGACCAAGATGAGCTTGAAGTGTGCGCAACGTGATGTG 1706
Qy 961 CAGACCTCGCGAGATGAGAGGCTGCGGCGCATCAACCCATGCGGAGAGCGGCGAG 1020
Db 1707 CAGACCTCGCGAGATGAGAGGCTGCGGCGCATCAACCCATGCGGAGAGCGGCGAG 1766
Qy 1021 CGCGGTGTGAGCGGAGTGCCTTTGCAAGACCGGACGATGTGTGTTGATGCGGCG 1080
Db 1767 CGTGGCGTTGAGCGGAGTGCCTTTGCAAGACCGGAGTGTGTGTTGATGCGGCG 1826
Qy 1081 TCGGCGCAATTCCTGTGGGCGGCTGATACCACTACCAATGATGCGCTGATGAC 1140
Db 1827 ACGGCGCAATTCCTGTGGGCGGCTGATACCACTACCAATGATGCGCTGATGAC 1886

Qy 1141 GAGACCGGCTTGTGAGGAGTGAAGAGTGGTGTGAAGAGCTGACGTTGATAT 1200
Db 1887 GAAACCGGTGTGTGACCGGTAAAGATATCTCTTAAAGATCTGGACACCGATC 1946
Qy 1201 GAGCTGTCCCGACCTTCTGAGTGGGCGGAGTGTGTGAGCGGACGTAACCGGAC 1260
Db 1947 CGCATTTGCCGACATTTCTTGGTGAAGCGGAGTGGCGGATCTTGAACCCGAT 2006
Qy 1261 ACCGCGATTTACTTCTTGGCGGTGAAGATCCCTGTACAGTATATATGCGCGTATCA 1320
Db 2007 AGCGGATCTACTTCTTCTTCCCTTAACCAACGCTGTGGGATTTGGGAGAGTGA 2066
Qy 1321 GAGTTTACCGGCTGACGCTTATTAACACGAGCGGACCGCAAACTGCGCGCGCTTT 1380
Db 2067 GAGTTTACCGGATGAGGCTTATCAACACGAGCGGACCTTACTGCTTGGCGGAA 2126
Qy 1381 GAAATATGAGCGCGGATGAGCGGATGATATCAACACGAGCGGACCTTGTGCGG 1440
Db 2127 GAAATATGAGCGCGGATGAGCGGATGATATCAACACGAGCGGACCTTGTGCGG 2186
Qy 1441 GAGCGCGCTGAGGAGGAGTACTGCGCGGTTTGTGAGCGGAGCGGAGTGTGTTAAC 1500
Db 2187 GAGCGCTGAGGAGGAGTACTGCGCGGTTTGTGAGCGGAGCGGAGTGTGTTAAC 2246
Qy 1501 GCGGAGCGACCGCTATTTCCGTGCTTCAAGCAAGAAACCGGAGACTTTGTGCGAG 1560
Db 2247 GCGGAGCGACCGCTATTTCCGTGCTTCAAGCAAGAAACCGGAGACTTTGTGCGAG 2306
Qy 1561 GCGGCTCTTGGAGGAGTGGGAGCGGAGGAGTCAAGCTAGCTGAGTGTGAGCGG 1620
Db 2307 ACCGCTGTGGAGGAGTGGGAGCGGAGGAGTCAAGCTAGCTGAGTGTGAGCGG 2366
Qy 1621 TATATCGCATCGGTGCGGCGGCTGTGACCTATGAGCAAGATTAACCGCGCTG 1680
Db 2367 TATATCGCATCGGTGCGGCGGCTGTGACCTATGAGCAAGATTAACCGCGCTG 2426
Qy 1681 GAGGCAATCGATTCGATCTCGGTGCTGATGAGCAATCTATGCTTTGCACTGCG 1740
Db 2427 GCGACCATCGATTCGATCTCGGTGCTGATGAGCAATCTATGCTTTGCACTGCG 2486

RESULT 5
US-10-162-713-3
; Sequence 3, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Soo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; APPLICANT: Yum, Do-Young
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method
; FILE REFERENCE: 1533 3740000
; CURRENT APPLICATION NUMBER: US/10/162,713
; CURRENT FILING DATE: 2002-06-06
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1740
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-3

Query Match 60.5%; Score 1052; DB 15; Length 1740;
Best Local Similarity 75.7%; Pred. No. 6.4e-303;
Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;
Qy 1 ATGAACCCCAACAGCTGCTTGCACACGCGCGCGCTGATGCTTACCGCGCGCCGCC 60


```

Db      1 ATGAAAGCGAAGCTTTTCTGTTTGCAGGCGTTCGCTTGCAGACTACAGCACAATT 60
Qy      61 GCATTCGCGAGGTAACCCCGATTACGATGAATCTGCGCAACCCGCGCTGTGTGA 120
Db      61 GCGCTTCTGATGAGACCCCGTCAACGAGAGCTGTGGCAAAACCCGCGCGCGGAGAA 120
Qy      121 TGAATTAACACGCGCGCAACCAAGAAAATATCGCACTGCGCGCTGACCAATCACT 180
Db      121 TGAATTAACACGCGCGCAACCAAGAAAATATCGCACTGCGCGCTGACCAATCACT 180
Qy      181 GCGCAACACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db      181 CCGAAGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 240
Qy      241 CAGGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db      241 CAGGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Qy      301 CAGGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db      301 CAGGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy      361 GTGCGCAACGCTAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db      361 GTGCGCAACGCTAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Qy      421 CTGATTTACGCTCATGCGCAACCACTGATGATGATGATGATGATGATGATGATGATGAT 480
Db      421 CTGATTTACGCTCATGCGCAACCACTGATGATGATGATGATGATGATGATGATGATGAT 480
Qy      481 GTATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Db      481 AGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Qy      541 ATTGTGCGCAATGCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db      541 ATTGTGCGCAATGCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Qy      588 ATGCTGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Db      588 ATGCTGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 597
Qy      601 TTATCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Db      601 TTATCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
Qy      658 AAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db      658 AAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Qy      721 GCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db      721 GCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Qy      781 GCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Db      781 GCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
Qy      841 AACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db      841 AACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Qy      898 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 957
Db      898 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 957
Qy      961 CAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db      961 CAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Qy      988 CAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1017
Db      988 CAGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1017
Qy      1021 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db      1021 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Qy      1077 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1077
Db      1077 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1077
Qy      1140 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db      1140 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140

```

```

Db      1078 ACGGCGCAATTCCTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
Qy      1141 GAGACCGGCGCTTGTGACGCGTGAACGAGAGATGCGGTGCTGAAGAGCTGACGTTGAAT 1200
Db      1138 GAAACCGGCGCTTGTGACGCGTGAACGAGAGATGCGGTGCTGAAGAGCTGACGTTGAAT 1197
Qy      1201 GACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
Db      1198 CGCATTTGCCGCAATTCCTGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1257
Qy      1261 ACCGCGCAATTCCTTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Db      1258 AGCGCGCAATTCCTTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1317
Qy      1321 GAGTTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Db      1318 GAGTTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1377
Qy      1381 GAAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Db      1378 GAAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1437
Qy      1441 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db      1438 GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1497
Qy      1501 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
Db      1498 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1557
Qy      1561 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620
Db      1558 ACCGCGCAATTCCTTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1617
Qy      1621 TATATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
Db      1618 TATATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1677
Qy      1681 GAGCGCAATTCCTTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1740
Db      1678 GAGCGCAATTCCTTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1737

```

RESULT 6

```

US-10-162-713-10
; Sequence 10, Application US/10162713
; Publication No. US20030228672A1
; GENERAL INFORMATION:
; APPLICANT: Choi, Eun-Sung
; APPLICANT: D'Elia, John
; APPLICANT: Kim, Hye-Sun
; APPLICANT: Kim, Mi-Seo
; APPLICANT: Lee, Jung Kee
; APPLICANT: Pan, Jae-Gu
; APPLICANT: Stoddard, Steven F.
; APPLICANT: Yum, Do-Young
; TITLE OF INVENTION: Sorbitol Dehydrogenases of Ketogulonigenium spp., Genes and Method
; TITLE OF INVENTION: Use Thereof
; FILE REFERENCE: 1533, 3740000
; CURRENT FILING DATE: 2002-06-06
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 3200
; TYPE: DNA
; ORGANISM: Ketogulonigenium sp.
US-10-162-713-10

```

Query Match 60.5%; Score 1052; DB 15; Length 3200;
 Best Local Similarity 75.7%; Pred. No. 7, 4e-303;
 Matches 1317; Conservative 0; Mismatches 420; Indels 3; Gaps 1;

QY 1 ATGATCCCAACAAGCTGCTTGGACACAGCGCGCGTGTATGTTGCTAACCGCGCCG 60
 Db 1241 ATGATGCGAAGTCTTTTCTGTTTGAAGGCTGTGCGCTTGCAAGCTAACGCAATT 1300
 QY 61 GCATTGCGGAGGTAAACCCGATTAACGATGAATCTGTCGCAACCCGCGCTGTGAA 120
 Db 1301 GCGCTTGCTGATGTGACCCCGTACCGACGAGCTCTGTGCAAAACCGCGCGCGGCA 1360
 QY 121 TGGATTACTAGGCGCGCAACAAAGAACTATGCGCACTCGCCCTGACCCAGTCACT 180
 Db 1361 TGGATAGCTACCGCGCAACAAAGAACTATGCGCACTCGCCCTGACCAAGTCAAG 1420
 QY 181 GCGGACAGCTGTGATGATGCAACTGATGTGGCCCGCGGAGTGAAGCGGCGGCGTA 240
 Db 1421 CCGGAGAGCTCGGTGAGCTGCAACTGTGTGGGCGCGCGCATGAACCGCGCAAGTC 1480
 QY 241 CAGGTACGCGGATGATCATGATGAGGTGATGTATCTGCAAAACCCGCGTGTGATC 300
 Db 1481 CAGGTACCTCGCTGATCCATGATGAGGTGATGTATCTGCGCAACCCGCGCATCATC 1540
 QY 301 CAGGCGGTGATGCGCAACAGCGGATCTGATCTGGGAAACCGCGCGCACTGCGCGC 360
 Db 1541 CAGGCGATCAGCTTAACCGCGGACCTGATCTGGGAAACCGCGCGCACTGCGCGC 1600
 QY 361 GTCGCGACGCTAAACGCGCAAGCGACCGGACGCGCGCTCGCCCTTTACGCGACGAC 420
 Db 1601 GTGGCAACGCTGAACAGCTTCGGTGAAGCGATCGCGGTATCGCGGTATCGGCAACAAC 1660
 QY 421 CTCTATTACGCTCATGAGCAACCATCTGATCGCGCTGATATGAGACGCGCGCAAGTC 480
 Db 1661 GTTACTCTGCTGCGGGAACAACCTGCTGCTGATGACGACCAACCGCGCAAGTC 1720
 QY 481 GTATTGATGCAACGTGATCGGCGGAGAGAGCGCTTGACAGTAAACAAGCGGCGCG 540
 Db 1721 AGCTGACGCTGACCGCGCGCAAGGAGAGATGTTTC---TAAGTGTGCGGCGCG 1777
 QY 541 ATTGTCGCAATGCGCTCATGCTGCGGCGGTTCACTGCGCAATATGCGCTATGATGAC 600
 Db 1778 ATGTCGCTAACCGCGGATGCTGCGCGGTTGCACTGCGCAATATGCGCTGCGCTGC 1837
 QY 601 TTATATCGGCGGACGATTCGCGGAGGAGAGTGTGCGGCAACCACTTATCCCG 660
 Db 1838 TTGCTTTGGGCGCATGACGCACTACCGGTGAAGAACTGTGGCGCACTTATCATCCG 1897
 QY 661 CAGCGCGGCAAGAGGCTGACGAGCTTGGGCGAATGATTCGAGGCGCGCTGATGACC 720
 Db 1898 AAGCGCGGTGAAGAGCGATGAACCTTGGGCGAACGATACGAAACCGCGCTGATGACC 1957
 QY 721 GGGCTCTGGGCTGATCACTATGATCCCGTGAACCACTTGTGTTTATGCTGACCC 780
 Db 1958 GGGCTCTGGGCGCAACACAGTACGACCCCGTCAACCACTGATTTCTACGATGTG 2017
 QY 781 GGGCTGGGCGCAGCTTCGAAACCAAGCGCGGACGCGCGGCGCAAGCTGTATGCGACC 840
 Db 2018 GCGGTGCGCGCGGTTGCGAAACCAAGCGGACCAACCGCGCGCAACATGATGCGCAAG 2077
 QY 841 AACACCCGCTTGGGCTGCGGTCCGACACGCGGAGATTTGTCTGGGTCAACGACCTTG 900
 Db 2078 AACACCCGCTTGGGCTGCGGTCCGACACCGGCGAATCTCTGGGTCAACCAACTCTG 2137
 QY 901 CCGGCGGACACTGGGACCAAGATGACGTTGAGATGATGTTGCGCAACGTCGATGAG 960
 Db 2138 CCGGCGGACACTGGGACCAAGATGACGTTGAGATGATGTTGCGCAATGTCGACGTC 2197
 QY 961 CAACCTTCGCGCGGATGAGAGGTCTGCGCGCAATCAACCCCAATGCGCGGACGCGCGAG 1020
 Db 2198 CAGCTTCGCGTGCATGACGCGCTGAATGATCAACCCCAACGCGCGCATGCGCGAG 2257
 QY 1021 CCGCGTGTGCTGAGCGGCTGCGCTTGAACCGCGGCAAGTGTGCTGATGAGCGCC 1080
 Db 2258 CGTGCGTGTGACCGCGGCTTCCGTGCAAAACCGGTACATGTGCGAGTTTCGACGCTGAA 2317
 QY 1081 TCGGCGAATTCCTGTGGGCGCGTGAATCAACCAATATGATGCGCTGATGAC 1140

Db 2318 ACGGCGAATTCCTGTGGCGCGGACACGCACTACCAAAACATGATGATGATGAC 2377
 QY 1141 GAGACCGGCTTGTGACGCTGAACGAGATCGGTGCTGAAAGCTGAGCTTGAATAT 1200
 Db 2378 GAAACCGGCTGTGACGCTGAATGAAGATATCACTTAAAGATCTGGAACCGGACTAC 2437
 QY 1201 GAGCTGCGCGGACCTTCCGTGGGTGGGCGGACCTGTGTGAGCGACGACTGAACCGGAC 1260
 Db 2438 GCGATTGCGCGCAATTTTGGGTGAGCGACCTGCGCGTGTGAGATCTTAAACCGGAT 2437
 QY 1261 ACCGCGATTATCTTGTGCGCGCTGAACAATGCTGTGATATATATGATGATCAA 1320
 Db 2498 AGCGGATCTACTTCACTTCCCTGAAACAGCCTGTGGGATTTGGCGGAGTCACTAA 2557
 QY 1321 GAGTTAGCGGCTGACGCTTATTAACACGAGCGGACCGCAAACTCGCGCGGCGCTT 1380
 Db 2558 GAGTTCAAGGCAATGAGAGCTTACACACGCGGACTTACTGCTTGGCGCGGAAAAA 2617
 QY 1381 GAAATATGCGCGCGCATGACGCGATGATATACGACCGGCGCACTTGTGTGCGCG 1440
 Db 2618 GAAATATGCGCGCGCATGACGCGATGATATACGACCGGCGCAAACTGTGTGCGTC 2677
 QY 1441 GAGCGCTTGTGCGGCACTACTCGCGCTTGTGTGACGCGGCGGCTGTGTCAAC 1500
 Db 2678 GAACTCTGCGGCTGCACTACTCGCGCTTGTGTGACGCGGCGGCTGTGTCAAC 2737
 QY 1501 GCGCGGACGCAACGCTATTTGCGGCTGCGCGGAGAAACCGCGGAGCTTGTGCGAG 1560
 Db 2738 GCGGCGACGATCGCTTCTTCTGCGGCTTCTGCGGAGAACTGTGCGAGACCTGTGCGAG 2797
 QY 1561 GCGGCTTGTGCGGCGGCGGACGCGGCGATGACCTTACGATGAGCTGTGAGCGGCTGCAA 1620
 Db 2798 ACCGCTGTGCGGAGCTGTGCTTCTGCGGCGCAACCGGTGTGAGATGAGAGCGGCTGCA 2857
 QY 1621 TATATCGGCACTGCGGCGGCGGCTGTGACCTATGCGCACTTGAACGCGCGCTGCGC 1680
 Db 2858 TACATCGGCACTGCGGCGGCGGCGGACCACTTACGCGGCGGCTGCGGCGCTGCGC 2917
 QY 1681 GAGCGAATGATTCGACCTCGGTGCGGATGATGATGATGATGATGATGATGATGATGAT 1740
 Db 2918 GAGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2977

RESULT 7
 US-10-369-493-34527
 ; Sequence 34527, Application US/10369493
 ; Publication No. US20030233675A1

GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; PRIOR FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 34527
 ; LENGTH: 2151
 ; TYPE: DNA
 ; ORGANISM: *Sphingomonas aromatica*ivorans
 US-10-369-493-34527

Query Match 8.7%; Score 152.2; DB 15; Length 2151;
 Best Local Similarity 52.6%; Pred. No. 9.4e-35;
 Matches 477; Conservative 0; Mismatches 388; Indels 42; Gaps 5;

QY 83 TTACGATGAATGCTGTGCGGAAACCGCGCGCTGTGAATGATTAATGACGCGCGCAACC 142

113 TCACGAGCGCCCTCATGCGCCGCGCCGAGCGGAGCGGATGCGACTATGCGCGGATT 172
QY 143 AAGAAATCTATCGGCACTCGCCCTGACCGCATCTCTGCGCAACGTTGCTAGTTC 202
Db 173 ATGCGCGAGCGCTTCTCCCTCATCCGATCATATGCGCAACGTCGCGCACTCG 232
QY 203 AACTGCTGAGGCGCGGAGATGAGCGCGGCGCGGACAG---GTACGCGCATGATCC 259
Db 233 GTCTGCTGCTGCTCATGACCTGAGACCGCGCGCGGAGAGCGACCGCCCTGATGC 292
QY 260 ATGATGCGCTGATGTATCTGCGCAACCCCGGATGATGATCGAGCGCTGATGCGCAA 319
Db 293 ATGACGCGCAGCTCATCTCATCTGACCGCGCTGATGATGATGAGCGGCTGATGCGCAA 352
QY 320 CAGGCGATCTGATCTGAGAACACCGCGCGCACTGCGCGCGCGCGCACTAAAGCC 379
Db 353 CCGGCGGCTCAAGTGTCTTACGATCCGAGTCCCGCGGAGACCTGTCGCGCTT 412
QY 380 AAGCGCA---CGAAGCGCGCGCTGCGCCCTTTACGCAAGACCTCATTTCACTCAT 436
Db 413 GCTGCGACGCGGTCAATCGCGCGCTGCGCTTATGCGCAAGAGCTTCTGCGGACGC 472
QY 437 GGGACAACTCTGATCTGCGCTGATGATGAGACCGCGCGCGCTGATCTGATGATGAGC 496
Db 473 TCGACGCTGCGCTGCTGCGCTGACCGAAGACCGAAAGTGTGTGCTGCAAGTGC 532
QY 497 GTGATGCGCGGAGAGACGCGCTTACCAATACACACGCGCGCGATGTCGCAATGCGC 556
Db 533 TCGTGGCCAAACGAGAGGACTACAC---ATCACGCGCGCGCGCGCGGTGAGAGGGA 589
QY 557 TCAATGCTGCGGCTTCCACTGCGCAATTTCCGCTTATGATGATCTTATCTGCGGAC 616
Db 590 AGGTTCTGATGCTGATGCGCGCGCTGCGAGTACAAAGCGCGCTTATGCTGCGCTACG 649
QY 617 ATTCCGCGACGCGTGAAGAGCTGTGCGCAACCACTTTATCCGCGCGCGCGCAAGAG 676
Db 650 ACGTCAACACCGCGCAAGAGTGTGAAGTTCACACACGCTCCCGGCAATCCAGCGGACG 709
QY 677 GTACGAGACTTGGGCAATGATTTGAGGCGCGCTGATGACCGGCG----- 723
Db 710 GATTCGAGAACAGAGCGATGAGAAACGCGCGCGCACTGCGCTGCGATGATGATGAGAC 769
QY 724 -----GTTGCGGCTCAGATCACTATGATCCCGGAGAACTTTGAT 766
Db 770 TCGCGCGGCGCGCGAGTGTGAGATTCATCTACCTTATGATCCCGCACTTCTGCTCC 829
QY 767 TCTATGCTGACCGCGGTGAGCCACGCTCGAAACCGACGCGCGCAAGCGCGGCGCA 826
Db 830 TGTTCGCAACCGCATGCGAGCATGAGAACCGCGCGCGAGCGCGGCGAGGAGCA 889
QY 827 CGCTGATGAGCAACACCGCGTTTGGCGTGTGCTCCGACACGCGCGAGATTTGCTGCG 886
Db 890 GCGCTCA---CACTGCTCATGCTGCGGTGATCCGATACCGCGCATATGCTGCG 946
QY 887 GTACAGAGACCTGCGCGCGCAACCTGAGAACCAAGATGCACTTTCAGATGATGCTG 946
Db 947 ATTTCAGAGAAACCGGAGAAACCGCTGAGATTTGACCTCGCGGACAGATCAAGCTG 1006
QY 947 CCAAGCT 953
Db 1007 CCGACCT 1013

RESULT 8

US-10-369-493-45426
; Sequence 45426, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.

APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 45426
; LENGTH: 1731
; TYPE: DNA
; ORGANISM: Rhodobacter capsulatus
US-10-369-493-45426

Query Match 5.7%; Score 99.4; DB 15; Length 1731;
Best Local Similarity 44.5%; Pred. No. 4.8e-19;
Matches 739; Conservative 0; Mismatches 881; Indels 41; Gaps 7;

QY 104 ACCGCGCGCTGTGATGATTAATACGCGCGCAACCAAGAAACTATGCGCACTGCG 163
Db 80 AGCAGACCGCGCGCGAGTGTGACGACGAGATGAGGCGCGCATCTGACCGGTTGAGCC 139
QY 164 CCGTGAACCGATGACGCGCGCAAGCTGTGATGATGATGATGATGATGATGATGATGAT 216
Db 140 CGCTTGCACTGATACCGGCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 139
QY 217 --CCGCGAGATGAGCG 274
Db 200 TGGCGGCGGAGAGACG 259
QY 275 ATCTGCGAAACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334
Db 260 ACATACCGGCTCTATCTGCGGCTTACCGATGATGATGATGATGATGATGATGATGAT 319
QY 335 GCGAAGACCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
Db 320 GCGAATGAGCG 376
QY 395 GCGCGCTGCGCTTTACGCGCAAGCTCTATTTCACTATGATGATGATGATGATGATGAT 454
Db 377 GCGCGCGCGCGATGATGAGGACATGATGATGATGATGATGATGATGATGATGATGAT 436
QY 455 CGCTGATGATGAGACG 514
Db 437 GCGTGAACCGGAGACCG 496
QY 497 GCTATTC---TACACCG 553
Db 515 GCTTGAACGATGACCAACG 574
QY 575 CCGTGAATGATGCG 634
Db 554 CCGGCGCGCGAATGCG 613
QY 635 AGCTGTGCGCAACCACTTTATCCGCGAGC-----CGGCGAAGAGGATGA 680
Db 614 TGTCTGACCG 673
QY 681 CGAATCTTGGGCAATATTTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
Db 674 CGATGACG 732
QY 741 CTATGATCCGATGACGACCTTGTGTATGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 800
Db 733 GCGACCTGCTTGGCG 792
QY 801 AACCGACGCG-----GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 852
Db 793 AACCG 852
QY 853 GCGGTGCTCCGACG 912

Db	853	GGCGCAACACTTTATGCGCGCTTCGCGCGGGGATCGAACCGGCGACGGGCGAGATCAAG	912
OY	913	TGGGACCAAGATGCAAGCTTCGAATGATGCTGCGCAACGTGCAATGTCACAACTTCGGCC	972
Db	913	TGGCACTTTCAGACCAACCCCGCGGAGGGCTGGGATTTTCAGAGGCGTAAACGAGGTGGT	972
OY	973	GAGATGAGAGGCTCTGCGGCGCATCAACCCCAATGGCGGACCGGCGGAACGCGCTGTGTG	1032
Db	973	GCGCTTTGTCGCAAGAGACGGCAACGAGGCTTTGGCCACCGCGCAACGGATTTCTTT	1032
OY	1033	AAGGTTGGCGCTTCGAAGCCGGGCGAGATGTGTGTTGAATGGGCGCTCGGGCGCAATTC	1092
Db	1033	TATGTGCTTCACCGCGCGGACGGCAAAATTCGCAACGCTGGCCCTTGTGTGAAGAACTC	1092
OY	1093	CTGTGGCGCGGTGATACCAACTCAACCAATATGATGGCTTCGATCGACGAGACGGGCTT	1152
Db	1093	AACCTGGGCGAAGGCGATCGATGAACCGGCGCTCGCATCTCAACGAGAGAAACCGGCCC	1152
OY	1153	GTGACGCTGAACGAGAGATGGCGGTGCTGAAAAGCTGGAAGTTGAATATGACGTCTGGCC	1212
Db	1153	GGTGCCCCCGATGCGCGGCGCGAACCGGCGCAAGGGCCAGAGCATTTTCCGTGCTTCG	1212
OY	1213	ACCTTCTCTGGGTGGGCGGACTGTGCTGACGCGCACTGAACCCGGAACCGGCAATTTAC	1272
Db	1213	---TTCCCTGGGCGGCGAAGAACTGGAATGCGATGCGCTCAAGCGAAGACCAACGCGGCTCTTC	1268
OY	1273	TTCTTGGCGCTGAACATATGCTCGTACCATATATATGGCGCTTGATCAAGAGTTTACGCG	1332
Db	1270	TATGTCCCTTCGAAAGAAATGGGGATGTGATATCTTGAAACGAGCCGATCAATTTACAAGAG	1322
OY	1333	CTTCACGCTCTATTAACACGACGGCGAACCGCAAACTCGCGCGGGCTTTGAATAATGGGC	1392
Db	1330	GAGCGGGCTTACCTGGGGCGGGGCTTCCAGATCAAGCCGCTTTCGAGAAATTAACTCGSC	1388
OY	1393	CGCATCGACGCGATTTGATATCAGCACCGGCGCACCTTGTGGTCGCGAGCGGCTTCG	1452
Db	1390	TGCTCTAAGGGGATCGATCCGAATACGGGCGGCTGAATATGGAAATACAAAGACGACGCG	1448
OY	1453	GCGAACACTACTGCCCGCTTTTGTTCACGCGCAGCGGCTGTGTGTTCAACGGCGGACCGAC	1512
Db	1450	CCGCTGTGGGGGGGGGTATAGCACCGCGGGCAAGCTGTGTTCAACGGGCAACCCCGAG	1509
OY	1513	CGCTATTTCCGTGCGCTTCAGCCAGGAAACCGGCGAGACTTTGTGGCAGCGCCGTTCTCG	1572
Db	1510	GCGCAGTTTCATGGCGCTGCGAGCGGAAACCGCGAAAGCTGTGTGTTTCAAGACGCGC	1568
OY	1573	AACGTCGCGACGGGGCAGCGCATCAGCTACAGATTGACGGCGTGCATATATGCGCATC	1632
Db	1570	TGCGGCAATCGTGGCGCAACCGCATCTCTGGGAACAGACGGCGAGCAATATGTCTCGATC	1629
OY	1633	GATGCGGCGGCTTCGACCTATATGACGCAATTGAACGCGCGCTGGCGGAGGCAATCGAT	1692
Db	1630	ATCTCGGCGCTGGGGCGCGCGGTCGCTTTTGGGCGCGCAAGTGGCCAAAGCGGTCAAC	1689
OY	1693	TGCAACTCGGTGCGTAATGCAATCTATGCTTTTGACATGCC	1733
Db	1690	TATCTGAACCAAGGCGCGACGCTTCGACCTTCGCGCTGCC	1730

```

, RESULT 9
, US-10-369-493-28294
, Sequence 28294, Application US/10369493
, Publication No. US2003023675A1
, GENERAL INFORMATION:
, APPLICANT: Cao, Yongwei
, APPLICANT: Hinkle, Gregory J.
, APPLICANT: Slater, Steven C.
, APPLICANT: Goldman, Barry S.
, APPLICANT: Chen, Xianfeng
, TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
, TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
, FILE REFERENCE: 38-10(52052)B
, CURRENT APPLICATION NUMBER: US/10/369,493

```

```

?
? CURRENT FILING DATE: 2003-02-28
? PRIOR APPLICATION NUMBER: US 60/360,038
? PRIORITY FILING DATE: 2002-02-21
? NUMBER OF SEQ. ID NOS: 47374
? SEQ. ID NO: 28294
? LENGTH: 1677
? TYPE: DNA
? ORGANISM: Burkholderia fungorum
? US-10-369-493-28294

```

Query Match	4.8%	Score 83.8	DB 15	Length 1677
Best Local Similarity	47.1%	Pred. No. 2.1e-14		
Matches 438; Conservative	0	Mismatches 467	Indels 24	Gaps 5

Qy	38	TGTTATTTGGTTAACCCGCGCCCGCGCATTTGGCGAGGTAAACCCGATTACCGATTAACTGC	97
Db	32	TGGTATTGGGATTCGGCCGAGCGCCGCGACGATTACCCAGCCGCTCACTATTGAGGGTTGA	91
Qy	98	TGGCGAACCCCGCCCGTGGTGAATGATTTAACTACGGCCGACACCAAGAAAATTATGCC	157
Db	92	CACGTGCGGAGAGCCACCCCGCGTGGCTCAAGTCTTACCGGACGTACAGCGCCAGCGC	151
Qy	158	ACGCGCCCTGACCCGAGATCACTGGCGACAAGTATTTGGCAGTTTCCACTGGTCTGGGCC	217
Db	152	ATTGCGCGCTCAAGCAGATTCGACACCGGGAAGTGAAGAAAGCTCAACACAGGTCTGAGCT	211
Qy	218	GGCGGATGGAAGCGGG-----GGCGGTACAGCTCAGCCCGATATCATATATGGCG	268
Db	212	ACAAAGTTTCGGCGGATCTGACGAGGGATTTGGAAGCCAGCCCGATGTTAAGCAGCT	271
Qy	269	TGATGATCTGGCAACCCCGGTGATCTGATCTCAGCGCGTGAATGCGCAACAGCGCATC	328
Db	272	ACCTTTTGTGACAGCGCGAAAGCAAGCTATATGATTTGATGGCGGACCGGCAAGC	331
Qy	329	TGATCTGGGAACCGCGCGCAACTGCGCCGCGCTGCGCAGCTAAAGCCCAAGGCGAC	388
Db	332	AGTTGTGGAAAGTTGCAACCGAAGCTCGGTGGGAATGTTTCAAGACGCAATGTTGCACG	391
Qy	389	G--CAAGCGCGCGCTGCGCCCTTTACGCGACGAGCTCTAATTTAGCTCAATGGACAAC	448
Db	392	TGATCAACCGCGTGTGGGCTGTACCGCAGAAATGTCTACGTGCGATGTGAGCGGCG	451
Qy	446	ATCTGTATGCGGTGGATATGAGAGACGGCCAGCGCTGATTTCCAGTGTGAACGTGATCGG	505
Db	452	ACGTGCTGCGCTGACCGCGCAGAGAGCGCGCTGCGCTGGCGCAACAGATGTTGAGC	511
Qy	506	GGGAACCGGCTTGAACAGTAAACACACGCGGCGCGATTTGTGCGCAATGAGCTCATCGTCG	565
Db	512	CGGGTCTGGCTA---CGCGTTCTGCTGGCGCGCTGCGCTGACGGCGCACTGGTGG	568
Qy	566	CGGGTTCACCTGCGCAATTCGCGCCCTATGGAATCTTTATCTGGGGGCAAGATTCGCGCA	625
Db	569	TGGGTACCGACCGCGCAATACGCGCGCGCGCGCTTTTATGCGCACTGAATCCGATTA	628
Qy	626	CGGTTAGGAGCTGTGCGCAACCATTTTATCCGACGCGGCGGAAGAGGTGACAGA	685
Db	629	ACGGCAATGTCTATGAGAAAGCTTCAAGGTGCCGCGCGCGGAGAAAGGCGCAGACA	688
Qy	686	CTTGGGCGCAATTTTCGAGGCGCGCTGATGACCGGCGTCTGGGGCTCAGATCACTCATG	745
Db	689	CGTGGCCGACCGCATGACGAACACGGCGGCGCGCGCTGTGCTACCGGCACTTAG	748
Qy	746	ATCCGTGACGAACCTTGTGTTCTATGGCTGACCGGCGTGGGCGCAGGCTCGAAAACC	805
Db	749	ACGCACAGCATGAAAGACCCCTTACTGGGGCGTGGGCAACCCCGGCGCTGCGCCGATC	808
Qy	806	AGCGCGGACGCGCGGCGGACGCTGTATGGCAACAACCGGCTTTGGCGATGCTCCG	865
Db	809	TGG-----ACGGGCGACAACCTCTAT---TCCGTTTGGCTACTGGCACTGGATCCGA	859
Qy	866	ACAACGGGAGATTCTTGGCGCTACCAAGACCTTCGCGCGGACCACTGGGACCAAGAT	925
Db	860	AAACCGGCGATTTCAATGAGCACTACAGTACAGAAACATGACACATGGAGACTTACGACG	919


```

QY 402 CGCCCTTACGCGACGAGCCTCTATTTCAGCTCATGGGACCAACCATCTGATCGCGTGA 461
DB 420 CGCGCTGTACAAAGGCAAGTGTTCGTCGCGGCTTATGACGCGCGCTGTCGCTCGA 479
QY 462 TATGAGACGGGCGCAGGTGCTATTTCATGTCGAAACGTGATTCGGGGGAGAACGGCTTGA 521
DB 480 TGGCGGACCGGCGACGAGGCTGGGAGATGACAGCGCTGATGACACAGACGACTCTCA 539
QY 522 CAGTAACACGACGGGCGCATTTGTTCGCAATGCGTCATGTCGCGGTTTCACCTGCA 581
DB 540 CACCATACCGGTCGCGCGGAGTTCACGCGCAAGTGTGATGGGACGCGCGCGG 599
QY 582 ATATTGCGCTTATGATGCTTTATCTCGGAGGACGATTCGCGACGGGTGAGAGCTGTG 641
DB 600 CAAATACGCGCGCGGCTTACGTCACCGCTTATGATCCGAGACAGGACAGCGCAAG 659
QY 642 GC-----GCAACCACTTTATCCGCGAGCGGCGGAG 674
DB 660 GCGCTGGTTACGGTTCGCGGCGATCCGAGCAAGCGCTTGAAGACGAGTCCATGAGAA 719
QY 675 GGGTACGAGACTTGGGCGCAATGATTTGAGGCGCGCTGATGACCG-----CGT 725
DB 720 GGGCGCGCAACCTGGGATCCGCGCGCAATGATGAGCTCAATGCGGTGCGCGCACCGC 779
QY 726 CTGGGGGTCAATGATCCTGATCCCGTGAAGCACTTGTGTTATGAGCTTCAGCGCGCT 785
DB 780 GTGGAGACCATGACTTTCGATCCGACTCACTTGTATGATGAGGACGCGGCAACG 839
QY 786 GGGCGGACGCTCGGAAACCGAGCGGCGACGCGGCGGCGGACGCTGATGAGCAACAC 845
DB 840 CTGCGCGTGAATCGGATCTGCGGACGCGCGGAGGAGCAATCTGATCTGCGCTG 899
QY 846 CGCCTTTCGCGTCCGTCGCGACGCGGCGGAGATTGTTCGCGTCAACACCTCGCGG 905
DB 900 GATGTCGCGCTGAATGCGGACGCGCAATGCTGCGACTATCAGAGAGCGCGCG 959
QY 906 CGACAACTGGGACCAAGATGCACTTTCGAGATGATGTCGCCAAGTGCATGTGCAAC 965
DB 960 CGACAACTGGGATTAATCTCGACCGGAGATGATCTCGGAGCATGGCATGAGCGG 1019
QY 966 CTGCGCGGAGATGAGGCTTGCCTGCG 992
DB 1020 CAAAGCGCGGCAAGTATCTCTGACGCG 1046

```

RESULT 12

```

US-10-369-493-31374
; Sequence 31374, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31374
; LENGTH: 1596
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-369-493-31374

```

Query Match 4.3%; Score 74.8; DB 15; Length 1596;
 Best Local Similarity 48.7%; Pred. No. 1e-11;
 Matches 264; Conservative 0; Mismatches 272; Indels 6; Gaps 2;

```

QY 248 CGCGGATGATCCATGATGCGTGTATCTGGCAACCCCGGTGATGTGATCCAGGCG 307
DB 182 CGCGGATGATTCGAACGCGGTGATGTTCAATACCACTCGTTTAATCACTGATGCGG 241
QY 308 TGGATGCGCAACAGGCGATCTGATCTTGGGAACACCGCGGCACTGCGCGCGCTCGCA 367
DB 242 TCGATGCGGTGACGCGCAAGAAATTTGACACTCAAGACAAAGATGGCGCGGTCACTA 301
QY 368 CGCTAAAGCCCAAGGCGACCGCAAGCGGCGTTCGCTTTAGCGGACGAGCTCTAAT 427
DB 302 CGTTTGTTCG---GGGCGCAACATCGCGGCGGTGCGCATATGCGCGGACCGTCTCA 358
QY 428 TCACTCATGGGACCAACCATTTGATTCGCGCTGGAATATGAGACGCGGCGAGTCTATTCG 487
DB 359 TGGGACCTTCGACGCGCAAGTCTGTCGCGCTGACGCAAAAACCGGACGCTGTGTGGC 418
QY 488 ATGTGCAACGTGATTCGGGCGAAGCGGCTTGACCAATGACCAACGCGGCGCATTTGCG 547
DB 419 AAGCGCAATTCGCGGATCCGATGAAGCTATTGGG---AAGCATGCGCGCGACCTGCG 475
QY 548 CCAATGCGGTATCGTTCGCGGTTTCCACTGCGCAATATTCGCGCTATGATGCTTTACT 607
DB 476 TCGACGCGCAAGTGTGATTCGCGGACGAAACGCGCGCAATATGATGATTCGCGTTCTCA 535
QY 608 CGGCGCAAGATTCGCGGCGGAGAGCTGTGCGGCAACCACTTTATCCGCGACCGG 667
DB 536 AGGCTTTCGACGCAACTCCGCTCAATTCCTGTAACCTTTATACGATTCGGAACCG 595
QY 668 GCGAAGAGGTGACGAGACTTGGGCGCAATGATTTTCAGGCGCGCTGATGACCGCGCT 727
DB 596 GCGAGGAAGTGTGTGGGCGCAAGAGACGCTACCGCGCGGAGGAGGCGCAACTCG 655
QY 728 GGGTCAATATCACTATATCCCGTGAAGAACTTGTGTTATGAGCTTCACCGCGCTG 787
DB 656 ACGCGAAGAAAGCAAGCTTCGCGAAGAGGCGGCACTTTACAAAGCGCTCGCGCG 715
QY 788 GC 789
DB 716 GC 717

```

RESULT 13

```

US-10-369-493-28616
; Sequence 28616, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 28616
; LENGTH: 2051
; TYPE: DNA
; ORGANISM: Burkholderia fungorum
US-10-369-493-28616

```

Query Match 4.3%; Score 74.8; DB 15; Length 2051;
 Best Local Similarity 48.7%; Pred. No. 1.1e-11;
 Matches 264; Conservative 0; Mismatches 272; Indels 6; Gaps 2;

QY 332 TCTGGGAACACCCGCGCCAACTGCCCCGCTGCGCACGCTTAAAGCCCAAGCGACCG-- 389
Db 340 TGTGGCGGTTGAGCCCGAGGTGACATGCAAGGTCAACCGCACGCTCTGCTGCACATG 399
QY 390 -CAAGCGCGGCGTCCGCCCTTTACGGCAAGCCCTATTTCAAGCTCATGGGACAACCATC 448
Db 400 TCAACCTGGCGTGGCGGTGGGGCGCGCAAGGCTTTCGTGCGCACTTCGATGGGTGA 459
QY 449 TGAATCGGCTGATATGAGAACGCGCCAGGTGCTATTCGATTCGAACTGATCGGGCG 508
Db 460 TGTACGCACTGATGCGCCGACCGCGGCTCGTCTGAAGAGCGACTTCATCGAAGACC 519
QY 509 AAGACGCTTGAACCAACCAACGCGGCGGATTTGCGCAATGGCGTATCGTCGG 568
Db 520 GAGCCCAAGGCGCAATTCACCGCGGCTCCGAAATCGGGCGGATGTCGTGTCATCG 579
QY 569 GTTCCACTGCAATATTCGCCCTATGATGCTTATCTCGGGCACGATTCGCGACG 628
Db 580 GCATGGCGCGCGGAGTACGAGCTGCGCGCTATGTCAGCGGCTGATCTGATACCG 639
QY 629 GTGAGAGCTGTGGCGCAACCACTTATCC-----GCAGC 664
Db 640 GAAAGCTGCGCTGGCTGTGGCACGTGGTGCCTCCGATCCCAAGCTGGGCGCGCAGAAA 699
QY 665 CGGCGCAAGAGGGTGAAGAGACTTGGGGCAATGATTCGAGCGCGCTGGATGACCGCG 724
Db 700 CGCCGAACTCGAAGCGCGCGCTGAAGACTTGGAGCCGAAACAGCCGCTGGGACATGCG 759
QY 725 TC-----TGGGGTCAAGTCACTATGATCCGTGACGAACTTGTGTTCTATG 772
Db 760 GCGGTGGACCGCGTGGATGCACTCACTACACCGGAAACCGGCTGTCTGTGCG 819
QY 773 GCTCGACCGGCGTGGGCGCCAGCGTCCGAACCGAGCGCGCACGCCGGCGGACGCTGT 832
Db 820 GCAACCGCAATGGCGGCTTTACGCCACGTCGAAGCGCTGCTGCGGGGGCGACACC 879
QY 833 ATGACCAACACCCGCTTTCGCGTGCATCCGACACGAGCGAGATTGTGCGCTCACCC 892
Db 880 TCTACCTCGCAGCGCTGTGCGCTGATCCGAAGACCGGCGGATGAAGTGGCACTACC 939
QY 893 AGACCTGCGCGCGACAACTTGGAGCCGAAGATTCAGATGATGTTGCCAAG 952
Db 940 AGGAAACGCGCGGCGCACTGGGACTTCAACGCCAACCGCAGATGATCTCACCCGCA 999
QY 953 TCGATGTCAACTCGCGCGGAGATGAGGGGCTGCGCGC 992
Db 1000 TGAAGATCGACGCGGAGACCGGCCAAGTCTGCTCCACGC 1039

Search completed: November 22, 2004, 20:44:06
Job time : 943 secs

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2004, 14:15:00 ; Search time 5784 Seconds
(without alignments)
10962.160 Million cell updates/sec

Title: US-10-802-682-4

Perfect score: 1740
1 atgaaccacacacgcgtctgct.....tcttgcaactgcgcagctaa 1740

Sequence: 1 atgaaccacacacgcgtctgct.....tcttgcaactgcgcagctaa 1740

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues
Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	58.8	3.4	941	BZ564301	pacs2-164
C 2	57.8	3.3	751	BZ561554	pacs2-164
C 3	57.2	3.3	580	AZ935093	Bj_Ba000
C 4	54.6	3.1	1014	BZ561555	pacs2-164
C 5	51.8	3.0	517	CA298989	SCMCLV103
C 6	51.8	3.0	742	CA272117	SCVPLB208
C 7	51.6	3.0	858	BZ56371	pacs2-164
C 8	51.4	2.9	925	CNS0091P	AL053013
C 9	50.2	2.9	658	CA133332	SCCORL103
C 10	50.2	2.8	701	CA130595	SCCORL103
C 11	48.6	2.8	647	CA150641	SCBFR2204
C 12	46.2	2.7	791	BH378255	AG-ND-142
C 13	46.2	2.7	815	BH400384	AG-ND-160
C 14	46	2.6	639	CA107253	SCRFH100
C 15	46	2.6	709	CL659599	PR10134C
C 16	45.6	2.6	298	AZ577200	04903 Sho
C 17	45.6	2.6	723	BZ549537	pacs1-60
C 18	45.4	2.6	925	CNS0091P	AL053013
C 19	45.2	2.6	659	CA131657	SCBGR104
C 20	45.2	2.6	1191	BZ555577	pacs1-60
C 21	45	2.6	640	CA270833	SCROB206
C 22	44.4	2.6	783	CC729915	OGUCO11TH
C 23	44.4	2.6	845	CC659386	OGUHX61TH
C 24	44.4	2.6	856	CG325844	OGM3074TH

C 25	44.4	2.6	861	9	CC659397	OGUHX61TV
C 26	44.4	2.6	862	9	CG325853	OGM3074TV
C 27	43.8	2.5	650	6	CA244227	SCSEFL408
C 28	43.6	2.5	559	4	BG739469	EM1_81_CO
C 29	43.6	2.5	578	2	BE592327	MS1_93_C1
C 30	43.6	2.5	613	2	AW678546	MS1_16_F0
C 31	43.6	2.5	625	4	BG739296	EM1_81_CO
C 32	43.6	2.5	681	2	BE593953	MS1_104_D
C 33	43.4	2.5	763	6	CA212935	SCCCL7C0
C 34	43.4	2.5	935	9	CNS006XK	AL066051
C 35	43.2	2.5	690	6	CD448001	EL01N0203
C 36	43.2	2.5	696	6	CD866097	AZ02_102J
C 37	43.2	2.5	728	9	CC672860	OGBPP74TV
C 38	43.2	2.5	795	9	CC626234	OGM058TV
C 39	43.2	2.5	821	9	CC672856	OGBPP74TH
C 40	43.2	2.5	839	9	CNS004NB	AL054280
C 41	43.2	2.5	889	9	CG334632	OGOCRA7TV
C 42	43.2	2.5	932	9	CNS0072Q	AL066742
C 43	43.2	2.5	1007	9	CC673577	OGUM38TV
C 44	43	2.5	718	5	BH294429	603605349
C 45	43	2.5	793	9	CG234022	OGZAP76TH

ALIGNMENTS

RESULT 1
BZ564301/c 941 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_4680.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_4680, genomic survey sequence.
ACCESSION BZ564301
VERSION BZ564301.1 GI:27189813
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE 1 (bases 1 to 941)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,T.L., Kaul,R. and Oiser,M.V.
TITLE Whole-genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: cgraymond@u.washington.edu
Class: shotgun.
FEATURES
source location/Qualifiers
1..941
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_4680"
/clone_lib="pacs2-164"
/note="Clinical isolate 2-164 Whole genomic shotgun
library."
ORIGIN
Query Match 3.4%, Score 58.8; DB 8; Length 941;
Best Local Similarity 46.4%; Pred. No. 0.00039;
Matches 192; Conservative 0; Mismatches 222; Indels 0; Gaps 0;
OY 1202 ACGTGTGCGGACCTTCTGTGGTGGCGGACGTGCTGTCAGCCGACGTAAACCCGACACA 1261
DB 490 AAGTATCACCCCGTCTCTCGGCGGACAGAACTGAAACCCGATGCGCTTACGCGGACGACA 431
OY 1262 CCGGATTACTTCTTGGCGGCTGAACAATGCTGCTAGATATATATGCGCGCTGATCAAG 1321

Db 430 CCGGCTGTTCTTACGTCGCGCCCAACCACTGAGAGAAACTACTGACCGAGAGGTCA 371
 QY 1322 AGTTTACGGCGCTGACGCTGTATTAACAACAAGCGCAACCCGAAACTCGCGCGCTTGG 1381
 Db 370 GCTATACGAAGGCGACCGCTTACCTTGGCATGGGCTTCCGATCAACGCGCATGTATGACG 311
 QY 1382 AAAATATGGGCGCGCATCGACGCGATTATATGACGACCGGCGCACTTGTGTGGCGG 1441
 Db 310 ACCACGTGGGACCTGCGCGCATGAGCCCGTACGCGGCAAGGTGTGTGGAAACA 251
 QY 1442 AGCGCCCTGGGCGAACTACTGCGCCGCTTTGTGACGCGAGCGGTGTGTGTCAACG 1501
 Db 250 AGGAACACCTGCGCTGCGGCGCGGCTGCTGCGCCACCGCGCAACCTGTGTGTACCG 191
 QY 1502 GCGGAGCCGACCGCTATTTCCGTCGCTCAGCGAGAAACCGCGGAGACTTGTGTGCGAG 1561
 Db 190 GCAACCGGCGAGGCTTCTTCAAGGCTTTCAGACCGCAAGCGGCAAGAGCTGTGAAAT 131
 QY 1562 CCGCTTTGCGACGCTCGCGACGCGGCGACGCGATCAAGTTCAGTTCAGCGCG 1615
 Db 130 TCCAGACCGGCGACGCGCATGCTCGCCACCGATCACTGGGAAAGACGCGGG 77

RESULT 2
 BZ561554 751 bp DNA linear GSS 17-DEC-2002
 LOCUS pacs2-164_3321.x1 pacs2-164 Pseudomonas aeruginosa genomic clone
 DEFINITION pacs2-164_3321, genomic survey sequence.
 ACCESSION BZ561554
 VERSION BZ561554.1 GI:27181766
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.

REFERENCE
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.
 TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
 JOURNAL J. Bacteriol. (2002) In press
 COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.
 Location/Qualifiers

FEATURES
 source
 1..751
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="2-164"
 /db_xref="taxon:287"
 /clone_1ib="pacs2-164_3321"
 /note="clinal isolate 2-164 whole genomic shotgun library."

ORIGIN
 Query Match 3.3%; Score 57.8; DB 8; Length 751;
 Best Local Similarity 49.9%; Pred. No. 0.0008;
 Matches 208; Conservative 0; Mismatches 197; Indels 12; Gaps 2;

QY 161 CCGCCCTGACCGCATCTGCGGACGATGTTGTGATTCGACATGCTGTGCGC----- 215
 Db 81 CCGCGCTCAAGCAGGCTCAAGCGCGACGATGTTCAAGCTGACCCCGGCTGTGTACT 140
 QY 216 ----CCGCGGATGAGAGCGCGGCGCTACAGGTCAAGCGCATGATCATGATGCGGTA 271
 Db 141 CTTTCGCGACGAGAAAGCGCGCGCAAGAGTCCAGGCGATGTCAGCGAGCGGTGA 200

QY 272 TGTATCTGCAAAACCCGCGTATGTATCCAGCGCTGATGCGCAACAGCGCATCTGA 331
 Db 201 TCTACGTGACGGCATCTTATCTCGCGGCTTTCGCTCGAGAGCAAGACCGGCAAGCGCC 260
 QY 332 TCTGGGAACACCGCGCAACTGTCGCGCGTGGCGCAAGCTAAAGCCGCAAGGACCGCA 391
 Db 261 TGTGACCTTCAACACACCGCGCTGCGGACGATCCGCGCTGTCTC--GACGTGATCA 317
 QY 392 AGCGGCGCTGCGCCCTTATGACGACGACCTCTATTTCAGCTGATGAGCAACCATCTGA 451
 Db 318 ACCGCGCGCGCGCATTTATGCGGACAAGGTCTTCTTCGCACTCCGACCGCTCGGTGG 377
 QY 452 TCGGCTGATATGAGACGCGGCGAGTCTGATTCGATGTCGACGCTGATCGGCGAAG 511
 Db 378 TGGGCTGAAACAAGAACACCGCGAGGTGTGTGAGAAAGATTTCGCGCATCAAGCGG 437
 QY 512 ACCGCTTGACAGTAAACACGAGGCGGCGATGTCGCAATGAGCGTCATGTCGCGG 568
 Db 438 CCGGCTACACATGATCTGCGCGCGCGCCACCTCTGTAAGAGCGCAAGACCGGCAAG 494

RESULT 3
 A2935093 580 bp DNA linear GSS 24-APR-2001
 LOCUS BJ_Ba0003G14r.B. japonicum BAC library Bradyrhizobium japonicum
 DEFINITION BJ_Ba0003G14r.B. japonicum BAC library Bradyrhizobium japonicum genomic, genomic survey sequence.
 ACCESSION A2935093
 VERSION A2935093.1 GI:13777242
 KEYWORDS GSS.
 SOURCE Bradyrhizobium japonicum
 ORGANISM Bradyrhizobium japonicum
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE
 AUTHORS Tomkins, J.P., Wood, T.C., Stacey, M.G., Loh, J.T., Judd, A., Goicoechea, J.L., Stacey, G., Sadowsky, M.J. and Wing, R.A.
 TITLE A marker-dense, sequence-ready map of the Bradyrhizobium japonicum Genome Res. 11 (8), 1434-1440 (2001)
 JOURNAL Genome Res. 11 (8), 1434-1440 (2001)
 MEDLINE 21376150
 PUBMED 11483585
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7268
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Class: BAC ends
 High quality sequence stop: 541.
 Location/Qualifiers

FEATURES
 source
 1..580
 /organism="Bradyrhizobium japonicum"
 /mol_type="genomic DNA"
 /strain="USDAL10"
 /db_xref="taxon:375"
 /lab_host="E. coli"
 /clone_1ib="B. japonicum BAC library"
 /note="Vector: pindigo536; site_1: HindIII"

ORIGIN
 Query Match 3.3%; Score 57.2; DB 8; Length 580;
 Best Local Similarity 50.8%; Pred. No. 0.00095;
 Matches 193; Conservative 0; Mismatches 178; Indels 9; Gaps 2;

QY 118 GAATGGATTACTAGCGGCGCGCAACGAAGAACTATCGCCCTGAGCCGAGATC 177
 Db 383 GACTGGCTGACTTATTCACCGGCTCTTAATAGTCTTATACATCAAGTCCCTTGGCAGATC 324
 QY 178 ACTGCGCAACAGTGTGCTAGTTGCAACTGCTGTGCGC-----CGCGGATGAGGCG 231
 Db 323 AATACGAGCAATGTGGTAACTTAAGCGGTGCGGATCATATTCGGGAGCGGTCCACC 264

```

QY 232 GGGGCGGTACAGCTCAGCCCGATGATCCATGATGCGCTGATGTAATCTTGGCAAAACCCCGGT 231
Db 263 CGTGGCTTGCAAGTCAAGCCGCTTGCGCCGACGCGCTGCTCTACTACAGAGGCTCTTAC 204
QY 292 GATGTATCCAGGCGCTGATGTCGCAACAGAGCATGATCTGTGGAACACCGCCGCA 351
Db 203 AGTGAAGTGTGGGCTGAACGCGGACAGGAGGTGATCTGTCTCTTATTTCCCGAA 144
QY 352 CTGCCC--GCGTGCACGCTTAACGCGCCACGCAACGCGCGCGTGCCTT 408
Db 143 CTTCACGAAGCCCTGATGCGCGGACAGCGACTCTCCCTATATATCGCGCGCTGCCTC 84
QY 409 TACGCAACGAGCCCTTATTTTCACTCATGAGCAACCATGATGCGGCTGTATATGAG 468
Db 83 GGAAGAGGCAAGTCTTACCTGCGACATGATGAGCCGCTCATGCGCTTGAATGAG 24
QY 469 ACGGCGCAGGTCTATTGCA 488
Db 23 ACTGAAAGGTCCGCTGGA 4

RESULT 4
BZ561555/c 1014 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_3321.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_3321, genomic survey sequence.
ACCESSION BZ561555
VERSION BZ561555.1 GI:27181769
KEYWORDS GSS.
ORGANISM Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1014)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES
Source Location/Qualifiers
1..1014
/mol_type="Pseudomonas aeruginosa"
/db_xref="taxon:2164"
/strain="2-164"
/db_xref="taxon:287"
/clone_lib="pacs2-164_3321"
/clone_id="pacs2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."
ORIGIN
Query Match 3.1%; Score 54.6; DB 8; Length 1014;
Best Local Similarity 48.8%; Pred. No. 0.0048;
Matches 147; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 1381 GAAATATGAGCCGATGACGCGATTGATATGACACCGGCGCACTTGTGTCGCG 1440
Db 395 GACACGTCGCGAGCTGCGGCGCATGACCGGTCAGCGGCAAGTGTCTGGAAAC 336
QY 1441 GAGGCGCTTGCGGGAATCTAGCCCGTTTGTGACGCGCAGCGGTGTGTTCAAC 1500
Db 335 AAGGAACCTTGCGCTGCGCGCGGGGTCTGGCCACCGCGCAACCTTGAGTTCAAC 276
QY 1501 GCGGAGCAGCAGCTATTTCGTCGCTCCTCAGCCAGGAACCGGCAACTTGTGGCAG 1560

```

```

Db 275 GGACCGGAGAGGCTACTTCAAGGCTTTCGACGCGGAAGGCGGAAGAGCTGTGAAA 216
QY 1561 GCCCGTCTTGCCAGCGTGCAGCGGAGGAGCGCATACAGTTCAGTTCGCGCTGCA 1620
Db 215 TTCAGAACCGGAGGAGGAGTCTCTCGGACCGCATCACTGGGAACAGGACGCGACG 156
QY 1621 TATATGCCATGATGTCGCGCGCGCTGTACCTATGCAAGCAATTGAACGCGCGCTG 1680
Db 155 TACCTCGAGTAGAAGCTGCGCTACGGAACCGCGTGTGTGGGCGGCGACATGCGC 96
QY 1681 G 1681
Db 95 G 95

RESULT 5
CA299889 517 bp mRNA linear EST 26-SEP-2003
LOCUS SCMCIV1031F09.g LV1 Saccharum officinarum cDNA clone SCMCIV1031F09
DEFINITION 5', mRNA sequence.
ACCESSION CA299889
VERSION CA299889.1 GI:36073038
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Saccharum officinarum
Bakeryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
complex.
REFERENCE 1 (bases 1 to 517)
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Centro de Biologia Molecular e Engenhariaia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parud@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 031 row: F column: 09
Seq primer: 77 Promoter Primer.
FEATURES
Source Location/Qualifiers
1..517
/mol_type="Saccharum officinarum"
/db_xref="taxon:4547"
/clone_lib="SCMCIV1031F09"
/lab_host="DH10B"
/clone_id="LV1"
/note="Organ: Etiolated leaves from in vitro grown
seedlings; Vector: pSPORT1; Site_1: SalI; Site_2: NotI; An
unidirectional cDNA library generated from [Etiolated
leaves from in vitro grown seedlings]. cDNA was prepared
from polyA+ mRNA using SuperScript Plasmid System Kit
(invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments siting
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://succest.lad.ic.unicamp.br/public"
ORIGIN
Query Match 3.0%; Score 51.8; DB 6; Length 517;
Best Local Similarity 53.1%; Pred. No. 0.023;
Matches 110; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
QY 814 ACGCGGCGGACAGCTATGACCAACACCGCTTGGCGTGCCTCCGACACGGGC 873

```

Db 261 ACAGACGACAGGGGCGGCGCTGACAGGACGCGCGGTGAGTACGAGCTTTCGTGCGGCC 320

QY 874 GAAATGTCTGGGTGTACCAAGACCTTCGCCCGGGAAGTGGACCAAGATGCAAGCTTC 933

Db 321 GAACTGGGATGTGACCGCGAGAGCCCGGACGCGGACCACTGGGGCTTCCCAAGGTGGCC 380

QY 934 GAGATGATGTGCTCCCAACGTGATGTGCAACCTTCGCCCGAGATGAGGGGTGCGCGCC 993

Db 381 GCGCTGACGGGCGCTCAACAACGCGGACCAACCGCGGTATGACATCAAGGCGCTGCTCGCC 440

QY 994 ATCAACCCCAATGCGCGGACGCGCGAG 1020

Db 441 ATGAGGCTGTGTGCGCGCTGCGCGCG 467

RESULT 6
CA272117 742 bp mRNA linear EST 26-SEP-2003
LOCUS CA272117
DEFINITION SCVPLB2086D05.g LB2 Saccharum officinarum cDNA clone SCVPLB2086D05
5', mRNA sequence.

ACCESSION CA272117 GI:35983014
VERSION CA272117
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum

REFERENCE
AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parvud@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
plate: 086 row: D column: 05
Seq primer: T7 Promoter Primer.

FEATURES
source
Location/Qualifiers
1..742
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCVPLB2086D05"
/lab_host="DH10B"
/clone_id="LB2"
/note="Organ: Lateral buds from plants adult plants
growing in greenhouse; Vector: pSport1; Site: 1: Sal;
Site 2: NotI; An undirectional cDNA library generated
from [lateral buds from plants adult plants growing in
greenhouse]. cDNA was prepared from polyA+ mRNA using
SuperScript plasmid System Kit (Invitrogen). The
double-strand cDNAs were fractionated in a sepharose
CL-2B 40cm-columns and fragments sizing between 0.8 and
1.5 Kb were directionally cloned into the vector. Details
of each source of RNA and library construction can be
obtained at http://sucest.fed.unicamp.br/public"

ORIGIN

Query Match 3.0%; Score 51.8; DB 6; Length 742;
Best Local Similarity 52.0%; Pred. No. 0.025;
Matches 116; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

QY 814 ACAGCGGCGGACGCTGTATGACACACACCGCTTTCGGTGGGTCCCGACAGGCG 873
Db 147 ACAGCGTACAGGCGGCGGACGACGACGCGCGCTGTGACTACGCGCTTTCGCGGCC 206

QY 874 GAAATGTCTGGGTGTACCAAGACCTTCGCCCGGGAAGTGGACCAAGATGCAAGCTTC 933

Db 207 GAACTGGGATGTGACCGCGAGAGCCCGGACGCGGACCACTGGGGCTTCCCAAGGTGGCC 266

QY 934 GAGATGATGTGCTCCCAACGTGATGTGCAACCTTCGCCCGAGATGAGGGGTGCGCGCC 993

Db 267 GCGCTGACGGGCGCTCAACAACGCGGACCAACCGCGGTATGACATCAAGGCGCTGCTCGCC 326

QY 994 ATCAACCCCAATGCGCGGACGCGCGAGCGCGGTGTGCTGACGG 1036

Db 327 ATGAGGCTGTGTGCGCGCTGCGCGCG 369

RESULT 7
B2566371 858 bp DNA linear GSS 17-DEC-2002
LOCUS B2566371
DEFINITION* pacs2-164_6041.y2 pacs2-164 Pseudomonas aeruginosa genomic clone
pacs2-164_6041, genomic survey sequence.

ACCESSION B2566371 GI:27195788
VERSION B2566371
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa

REFERENCE
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
J. Bacteriol. (2002) In press
COMMENT Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source
Location/Qualifiers
1..858
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_6041"
/clone_id="pacs2-164"
/note="Clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN

Query Match 3.0%; Score 51.6; DB 8; Length 858;
Best Local Similarity 49.6%; Pred. No. 0.028;
Matches 132; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 1468 GTTGTGTGACGAGCGGCGGTGTGTCAACGCGCGGACCAACGCTATTTCCGTGCC 1527

Db 565 GTGTGGCCACACCGTCAACCTGTGTCAACGCGCACGACGACGCGCTACTTCAAGGCC 506

QY 1528 CTGACCCAGGAACCGCGAGACTTTGTGTGAGGCGCGTCTTGTGACGCGCGAGCGGG 1587

Db 505 TTGACCGGAGAGCGGAGAGAGCTGTGAAATTCCAGACCGGACGCGGATGCTTTCG 446

QY 1588 CAGCGATCACTACGATTGAGACGCGCGGTGCAATATATCCATCGGTGCGGGCGGTCTG 1647

Db 445 CCACCGATCACTCTGGAGAACAGAGCGGAGAGTACTCGGCGTGAACCGCTACGCGC 386

QY 1648 ACCTATGACGCAATTGAAGCGCCCGCTGCGCCGAGGCAATTCATTGACCTCGTGGT 1707

Db 385 GCGCGCGGCGCGCTGTGGGCGGAGACATGCGCTGACCTGACCGCGCGGTGGCCAGAGGC 326

QY 1708 AATGATCATATGCTTTGCACTGCC 1733

Db 325 GCCTACGGGCGCTCAACAACGCCGACACCCGCTGTACGACATCAAGGCGCTGCTGCC 384
 QY 994 ATCAACCCCAATGCGGCGGAGCGCCGCTGTGTGTAACGG 1036
 Db 385 ATGAGCGCTGGCGCGCGGATGCGCCGACGCGCGCGCGCTGG 427

RESULT 10
 CA190595 701 bp mRNA linear EST 24-SEP-2003
 LOCUS CA190595
 DEFINITION SCCCRT1C03609.g R11 Saccharum officinarum cDNA clone SCCCRT1C03609
 5', mRNA sequence.

ACCESSION CA190595
 VERSION CA190595.1 GI:35134909
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 701)
 AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
 TITLE The libraries that made SUCEST
 JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 COMMENT Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089

Email: parnuda@unicamp.br
 Clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bcccenter.fcav.unesp.br
 Plate: C03 row: G column: 09
 Seq primer: T7 Promoter Primer.

FEATURES
 source location/Qualifiers

1..701
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCCCT1C03609"
 /lab_host="DH10B"
 /clone_1lb="R11"
 /note="Organ: Root tips (0.3cm-long) from adult plants;
 Vector: pSport1; Site_1: SalI; Site_2: NotI; An
 unidirectional cDNA library generated from [root tips
 (0.3cm-long) from adult plants]. cDNA was prepared from
 polyA+ mRNA using Superscript Plasmid System Kit
 (Invitrogen). The double-strand cDNAs were fractionated
 in a sepharose CL-2B 40cm-columns and fragments sizing
 between 0.8 and 1.5 Kb were directionally cloned into the
 vector. Details of each source of RNA and library
 construction can be obtained at
 http://sucest.lad.ic.unicamp.br/public"

ORIGIN
 Query Match 2.8%; Score 50.2; DB 6; Length 701;
 Best Local Similarity 51.8%; Pred. No. 0.063;
 Matches 115; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 814 ACCGCCGGCGGCAACGCTGTATGCGACCAACCCGCTTTCGGTGTCCGACAGGAGC 873
 Db 205 ACGACGTCGACAGGGGGGGGAGGAGGAGCGCGCTGTGACTACGCGCTTTCGCGAGCC 264
 QY 874 GAGATTGTCTGGCGCTCAACCGACCTGCGCGGCGGACCACTGGGACCAAGATGCAAGTTTC 933
 Db 265 GAGCTGGGCAATGACACCGGAGAGCGCCGACGACCACTGGGCGCTTTCGCAAGGTGGCC 324
 QY 934 GAGATGATGTGCGCAACGTGATGTCACCTTCGCGCGAGATGAGGAGGTCTGCGCGCC 993

Db 325 GCCTACGGGCGCTCAACAACGCCGACACCCGCTGTACGACATCAAGGCGCTGCTGCC 384
 QY 994 ATCAACCCCAATGCGGCGGAGCGCCGCTGTGTGTAACGG 1036
 Db 385 ATGAGCGCTGGCGCGCGGATGCGCCGACGCGCGCGCGCTGG 427

RESULT 11
 CA150641 647 bp mRNA linear EST 24-SEP-2003
 LOCUS CA150641
 DEFINITION SCBFR2045E02.g R22 Saccharum officinarum cDNA clone SCBFR2045E02
 5', mRNA sequence.

ACCESSION CA150641
 VERSION CA150641.1 GI:35053896
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum

REFERENCE 1 (bases 1 to 647)
 AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
 TITLE The libraries that made SUCEST
 JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
 COMMENT Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica
 Universidade Estadual de Campinas
 Caixa Postal 6010, 13083-970, Campinas SP, Brazil
 Tel: 55 19 3788 1137
 Fax: 55 19 3788 1089

Email: parnuda@unicamp.br
 Clone distribution information can be found
 through the Brazilian Clone Collection Center (BCCC) at
 http://www.bcccenter.fcav.unesp.br
 Plate: 045 row: E column: 02
 Seq primer: T7 Promoter Primer.

FEATURES
 source location/Qualifiers

1..647
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCBFR2045E02"
 /lab_host="DH10B"
 /clone_1lb="R22"
 /note="Organ: Shoot-root transition zone from young plants
 (small insert library); Vector: pSport1; Site_1: SalI;
 Site_2: NotI; An unidirectional cDNA library generated
 from [shoot-root transition zone from young plants (small
 insert library)]. cDNA was prepared from polyA+ mRNA
 using Superscript Plasmid System Kit (Invitrogen). The
 double-strand cDNAs were fractionated in a sepharose
 CL-2B 40cm-columns and fragments sizing between 0.8 and
 1.5 Kb were directionally cloned into the vector. Details
 of each source of RNA and library construction can be
 obtained at http://sucest.lad.ic.unicamp.br/public"

ORIGIN
 Query Match 2.8%; Score 48.6; DB 6; Length 647;
 Best Local Similarity 51.1%; Pred. No. 0.16;
 Matches 114; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 814 ACCGCCGGCGGCAACGCTGTATGCGACCAACCCGCTTTCGGTGTCCGACAGGAGC 873
 Db 131 ACGACGTCGACAGGGGGGGGAGGAGGAGCGCGCTGTGACTACGCGCTTTCGCGAGCC 190
 QY 874 GAGATTGTCTGGCGCTCAACCGACCTGCGCGGCGGACCACTGGGACCAAGATGCAAGTTTC 933
 Db 191 GAGCTGGGCAATGACACCGGAGAGCGCCGACGACCACTGGGCGCTTTCGCAAGGTGGCC 250
 QY 934 GAGATGATGTGCGCAACGTGATGTCACCTTCGCGCGAGATGAGGAGGTCTGCGCGCC 993
 Db 251 GCCTGACGGGCGTCAACAACGCGGACGCGGTGTGACGACATCAAGGCGCTGCTGCC 310

QY 994 ATCAACCCCAATGCGGAGACGGCGGCGCTGTGCTACGG 1036
 Db 311 ATGAGGCGCTGATGCGCGATGCGCCGACGCGCGCGCTGG 353

RESULT 12
 BH378255/c
 LOCUS BH378255 791 bp DNA linear GSS 10-DEC-2001
 DEFINITION AG-ND-142L4.TR ND-TAM Anopheles gambiae genomic clone AG-ND-142L4,
 genomic survey sequence.
 ACCESSION BH378255
 VERSION BH378255.1 GI:17324397
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Anopheles gambiae (Arthropoda; Hexapoda; Insecta; Pterygota;
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 1 (bases 1 to 791)
 Ren,C., Huff,B.R., Carille,J.L., Black,K., Zhang,H.-B.,
 Gardner,M.J. and Collins,F.H.
 Construction of a BAC library and generation of BAC end
 sequence-tagged connectors for genome sequencing of the African
 malaria mosquito Anopheles gambiae
 Mol. Genet. Genomics 268 (6), 720-728 (2003)

JOURNAL
 MEDLINE
 PUBMED 12655398

COMMENT
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seg primer: M13 Rev
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..791
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-142L4"
 /clone_lib="ND-TAM"
 /note="Vector: pCBAC1; Site_1: HindIII"

ORIGIN
 Query Match 2.7%; Score 46.2; DB 8; Length 791;
 Best Local Similarity 64.5%; Pred. No. 0.7;
 Matches 69; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 110 CCGCTGTGAATGATTAATTAAGCGCCGCAACCAAGAAACTATGCCACTGCCCTGGA 169
 Db 194 CCCCCGCGATTGGAGCACTAGCGAACAACCCGCTGGCGATTACTGCGCTTGG 135

QY 170 CCGAGATCACTGCCGCAACGTTGTCTAGTTGCACTGCTGGGCC 216
 Db 134 ATCAATCACTCCGGCCAAAGTTGCCGATCTCACTGGCCCTGGACC 89

RESULT 13
 BH400384/c

LOCUS BH400384 815 bp DNA linear GSS 11-DEC-2001
 DEFINITION AG-ND-160D21.TR ND-TAM Anopheles gambiae genomic clone
 AG-ND-160D21, genomic survey sequence.
 ACCESSION BH400384
 VERSION BH400384.1 GI:17346600
 KEYWORDS GSS.
 SOURCE Anopheles gambiae (African malaria mosquito)
 ORGANISM Anopheles gambiae
 Anopheles gambiae (Arthropoda; Hexapoda; Insecta; Pterygota;
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 Anopheles.
 1 (bases 1 to 815)
 Ren,C., Huff,B.R., Carille,J.L., Black,K., Zhang,H.-B.,
 Gardner,M.J. and Collins,F.H.
 Construction of a BAC library and generation of BAC end
 sequence-tagged connectors for genome sequencing of the African
 malaria mosquito Anopheles gambiae
 Mol. Genet. Genomics 268 (6), 720-728 (2003)

JOURNAL
 MEDLINE
 PUBMED 12655398

COMMENT
 Other_GSSs: AG-ND-160D21.TF
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seg primer: M13 Rev
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..815
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-160D21"
 /clone_lib="ND-TAM"
 /note="Vector: pCBAC1; Site_1: HindIII"

ORIGIN
 Query Match 2.7%; Score 46.2; DB 8; Length 815;
 Best Local Similarity 64.5%; Pred. No. 0.7;
 Matches 69; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 110 CCGCTGTGAATGATTAATTAAGCGCCGCAACCAAGAAACTATGCCACTGCCCTGGA 169
 Db 191 CCCCCGCGATTGGAGCACTAGCGAACAACCCGCTGGCGATTACTGCGCTTGG 132

QY 170 CCGAGATCACTGCCGCAACGTTGTCTAGTTGCACTGCTGGGCC 216
 Db 131 ATCAATCACTCCGGCCAAAGTTGCCGATCTCACTGGCCCTGGACC 85

RESULT 14
 CA107253 639 bp mRNA linear EST 23-SEP-2003
 LOCUS SCRFR1007G10.g HR1 Saccharum officinarum cDNA clone SCRFR1007G10
 DEFINITION 5', mRNA sequence.
 ACCESSION CA107253
 VERSION CA107253.1 GI:34960560
 KEYWORDS EST.
 SOURCE Saccharum officinarum

